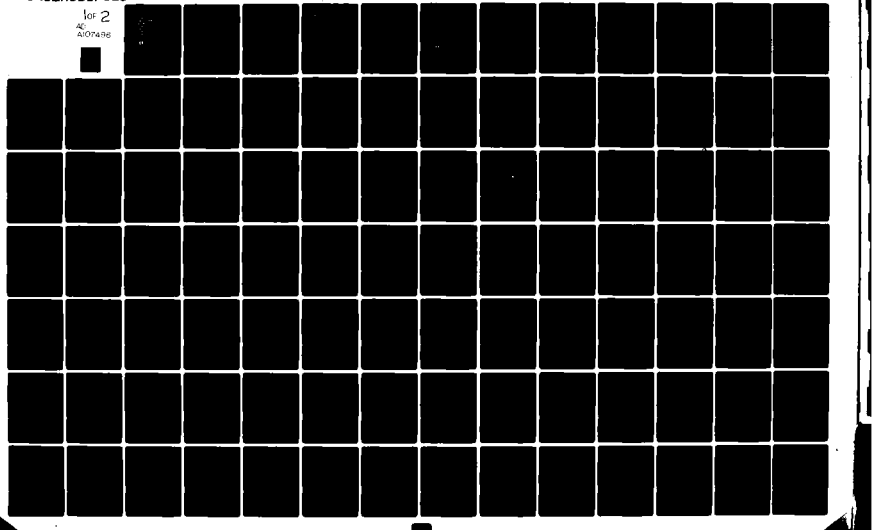
AD-A107 496 BOOZ-ALLEN AND HAMILTON INC BETHESDA MD F/G 5/1
THE RELATIONSHIP BETWEEN ADEQUACY OF TEST AND EVALUATION DATA A--ETC(U)
JUN 81 MDA903-80-C-0591
UNCLASSIFIED NL



THE RELATIONSHIP BETWEEN ADEQUACY OF TEST AND EVALUATION DATA AND DSARC DECISIONS

Prepared For
The Defense Systems Management College

Fort Belvoir, Virginia

BOOZ·ALLEN & HAMILTON INC.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A107496
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) The Relationship Between Adequacy of Test and Evaluation Data and DSARC Decisions	5. TYPE OF REPORT & PERIOD COVERED Report June 1981
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Defense Systems Management College	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Defense Systems Management College Fort Belvoir, VA 22060	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Defense Systems Management College, Info Center Bldg 205, Ft. Belvoir, Va 22060	12. REPORT DATE 1981
	13. NUMBER OF PAGES 88
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

~~UNLIMITED~~ UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

SEE ATTACHED SHEET

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

SEE ATTACHED SHEET

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 55 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

The objective of this research was to gain a better understanding of how the adequacy of test and evaluation (T&E) data affects defense system acquisition review council (DSARC) decisions. Specifically, the study focused on the adequacy of the T&E data presented at milestone reviews as a possible reason for conditional decisions. Conditional decisions are those that nominally approve entry into the next phase of acquisition, but also impose restrictions on the pace of development or rate of production. Such restrictions invariably increase the time and cost of the acquisition process.

This study showed that while most milestone II and III decisions are, in fact, made conditionally, the adequacy of T&E is not the predominant factor. However, when DSARC issues do arise which involve the quantity or quality of T&E data available for decision-making, it is generally because of a lack of communication between OSD and the services. Clarification of milestone review policies, early identification of T&E issues associated with the selected acquisition strategy, and vigorous use of the test and evaluation master plan (TEMP) are some of the ways this aspect of the acquisition process can be improved in the present environment of decentralized operations and centralized decision-making.

Key Words: Systems Acquisition Management
Acquisition Management
Acquisition
Acquisition Planning
Program Management
DSARC
Testing
Operational Test
Developmental Test
Testing
Test and Evaluation

INCL (2)

DEPARTMENT OF DEFENSE
DEFENSE SYSTEMS MANAGEMENT COLLEGE
FORT BELVOIR, VIRGINIA 22060

DRI-R

8 September 1981

SUBJECT: DSMC Contract Research Report: "The Relationship Between Adequacy of Test and Evaluation Data and DSARC Decisions".

SEE DISTRIBUTION (Inclosure 1)

1. Inclosed is a copy of the subject research report.

2. The objective of this effort was to gain a better understanding of how the adequacy of test and evaluation (T&E) data affects defense system acquisition review council (DSARC) decisions. Specifically, the study focused on the adequacy of the T&E data presented at milestone reviews as a possible reason for conditional decisions. Such restrictions frequently increase the time and cost of the acquisition process.

3. This research product is the result of a competitive award to Booz-Allen and Hamilton, Inc.; it documents all our work in data gathering, analysis, decision profiles (i.e. case examples) and includes findings, conclusions and recommendations on how test and evaluation data can be better integrated into the DSARC decision making process.

4. Research was essentially complete when, on April 30, 1981 the Deputy Secretary of Defense announced his approach to improving the acquisition process. The data and findings of this research bear directly on several of those decisions. This report should be useful in providing insight on T&E related issues, shortening the process, or improved planning.

5. Mr. John McKeown was research director. Should you have any inquiries, you may contact either him or the undersigned at:

Autovon	354-5783
Commercial	(703) 664-5783
Toll-Free	(800) 336-3095, X45783

Garcia E Morrow

GARCIA E. MORROW
LTC, U. S. Army
Dir., Research Directorate

DISTRIBUTION:

OUSDR&E:
  DDT&E (10)
  AM (5)

ARMY:
  USAOTEA (5)
  Dep Asst Sec (R&D) (SARDA)
  DCSRDA (DAMA-RAA)
  DCSRDA (DAMA-PPM)
  DARCOM (DRCDE)
  DARCOM (DRCQA)
  TRADOC (DCST&E)
  APRO (DARE)

NAVY:
  ASN (RE&S)
  CNO (OP-98,-983)
  COMOPTEVFOR (5)
  NAVMAT (MAT-09J,-08B,-08D)
  ONR (DARE)
  NCAR/NPGS

AIR FORCE:
  SAF/ALG
  SAF/ALP
  SAF/ALR
  AF/LEXY
  AF/LEYE
  AF/RDPT
  AF/RDQM
  AF/RDQM
  AF/RDXM
  AF/XOORE
  AFALD/CC
  AFBMRC (DARE)
  AFLC/LO
  AFLMC/CC
  AFSC/TEV
  AFSC/SD
  AFTEC (5)
  AUL/LD

MARINE CORPS:
  MCDEC
  MCOTEA (5)

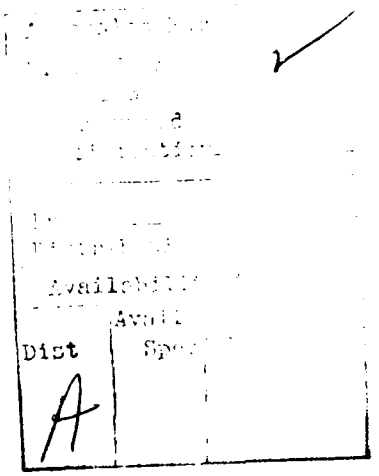
FAI
FAI Library

DTIC

Inclosure 1

Contract No.MDA-903-80-C-0591
June 1, 1981

THE RELATIONSHIP BETWEEN ADEQUACY OF TEST AND EVALUATION DATA AND DSARC DECISIONS



Prepared For
The Defense Systems
Management College

Fort Belvoir,
Virginia

BOOZ-ALLEN & HAMILTON INC.

FOREWORD

In a 1980 report of issues identified in studies of 21 major weapon systems, the General Accounting Office wrote:

> During weapon system development, the adequacy of testing is critical to assess and reduce program acquisition risks and to evaluate operational effectiveness and suitability. The purpose of testing is to minimize uncertainties that could adversely affect system effectiveness, cost, or availability for deployment. We found examples of testing not being comprehensive, rigorous, and complete.

The phrase, adequacy of testing, has, in a sense, become a battle cry for those who believe that the Department of Defense takes unwarranted risks in acquiring its weapon systems and moves precipitately toward greater and greater commitment to unproven, sometimes unworkable, designs. On the other hand, some, including the 1977 Defense Science Board, are concerned that current policies encourage overtesting and that the acquisition cycle may be lengthened because "...the independent test agencies will gradually gain increased influence and demand duplicate test activities."

The test agencies are sensitive to this criticism and believe they are executing their responsibilities as required not only by current policy but also by common sense. The user and developer have still different perspectives on adequacy of testing and "how much is enough?" Ultimately, the Office of the Secretary of Defense (OSD) decisionmakers must integrate the test and evaluation (T&E) data and other issues which surround each milestone review in order to make a rational decision.

In an overwhelming number of cases, these decisions have been made conditionally, i.e., they include qualifications that often delay or stretch out a program. This study focuses on the adequacy of T&E data presented to milestone reviews as a possible cause of the conditional decision syndrome. Reducing or eliminating this syndrome is one way to shorten the acquisition cycle and reduce costs; but, in order to do so, the underlying causes of inadequate T&E data or other factors that result in conditional decisions must first be understood.

No one model exists which characterizes all aspects of the decisionmaking process. This study analyzed individual program decisions in the context of changing policy and other external events. However, there is a wide variation in the way acquisition policy is applied to strategic as opposed to tactical or general purpose programs. Strategic and even tactical nuclear programs tend to operate in an expedient fashion and are less bound by established acquisition policy. Therefore, it may be inappropriate to generalize the application of the observed trends and other findings of this study to such programs.

This report is organized in six chapters. Chapters I and II provide background on the relationship of T&E to the DSARC decisionmaking process and define the study objectives and approach. Chapter III explores the way changes in acquisition and T&E policy and other external events merge to form a decision environment in which the milestone reviews are held. Chapter IV investigates the DSARC decisions for the period under study as a whole and analyzes trends and relationships. Decision profiles of six specific programs which involved questions of adequacy of T&E data at milestone III are provided in Chapter V. Finally, the findings of the study are integrated into conclusions and recommendations which are given in Chapter VI.

ADDENDUM TO FOREWORD

On 30 April 1981, during the final publication of this report, Deputy Secretary of Defense Carlucci announced 31 decisions on improving the acquisition process, see Exhibit F-1. These decisions, along with companion changes to the Planning, Programming and Budgeting System announced on 27 March 1981, represent the Reagan Administration's response to persistent problems in major systems acquisition as perceived by Congress and the GAO, the OSD staff, the Services, and individual Program Managers. While the body of this report has not been revised to reflect the new initiatives*, several initial observations can be made regarding the agreement between the new policies and procedures and the findings of this study.

* Consequently, references to "current policy" refer to the March 1980 revision to DODD 5000.1 and DODI 5000.2, and "the present" refers to the period prior to May 1981.

The recommendation of this report that OSD seek added funding flexibility during the critical transition from R&D to production in order to reduce the pressure on making premature production decisions is specifically addressed by the new initiatives. Also, the consolidation of OSD management attention at a new milestone II, "Program Go-ahead," which authorizes full-scale development and production and will occur somewhat later than the previous miletone II, is consistent with the finding of this study that a disproportionate level of management attention is now focused on milestone III and that a better balance between available information and the ability of the bureaucracy to influence program direction exists earlier in the program at around milestone II. It is not yet clear how the new, somewhat later, milestone review will be integrated into the traditional activities associated with full scale development and it remains to be seen whether this will result in substantive benefits in terms of T&E. It may be that an extensive, delayed milestone II review will permit some early developmental testing to be considered, but it is also possible that the change in emphasis will alter the role of operational test and evaluation in the decisionmaking process.

The new DSARC procedures also call for the production program review (old milestone III) to be delegated to the Service Secretary. Consequently, the recommendation of this study that the test and evaluation master plan (TEMP) be emphasized as a means of achieving structured communication between OSD and the Services is even more important than it was previously. A clear understanding of the measure of a system's readiness to proceed from R&D to production must exist in order for OSD to successfully delegate the rate production decision to the Services. The TEMP, if integrated into the T&E planning at "Requirements Validation" and "Program Go-Ahead," will be the primary mechanism for establishing and communicating these requirements. While one can only speculate on whether delegating the milestone III decision to the services will stand the test of possible Congressional, GAO or public pressure for OSD intervention, the TEMP could be the instrument which will help maintain the controlled decentralization that is now being sought.

Other new policy initiatives, such as providing adequate test hardware and realistically budgeting and fully funding programs, are consistent with the findings and conclusions of this study. However, the study also indicates that even increased funding and additional test hardware will not necessarily be effective in reducing conditional decisions if the critical test issues and test and evaluation plans are not identified and agreed to well in advance of the milestone review.

EXHIBIT P-1

INITIATIVES ON IMPROVING THE ACQUISITION PROCESS

On April 30, 1981, Deputy Secretary of Defense Frank Carlucci announced major changes both in the acquisition philosophy and the acquisition process as practiced by the new administration. Based on a 30 day assessment of the Defense acquisition system, the decisions address the major problems in system acquisition perceived by Congress and the GAO, the OSD staff, the Services and Program Managers. The major theme of the changes is to achieve enhanced readiness, reduced acquisition costs and shortened acquisition time through controlled decentralization. Implementation of the 31 decisions is presently underway.

1. Management Principles include improved long-range planning; greater delegation of responsibility, authority and accountability, emphasis on low-risk evolutionary alternatives; more economic production rates, realistic budgeting and full funding, improved readiness and sustainability; and strengthening the industrial base.

2. Preplanned Product Improvement should be used as a means of achieving performance growth.

3. Multiyear Procurement should be used, on a case-by-case basis, to reduce unit production costs.

4. Increased Program Stability in the Acquisition Process should be achieved by fully funding R&D and procurement in order to maintain the established baseline schedule.

5. Encourage Capital Investment to Enhance Productivity through legislative, contractual and other economic incentives.

6. Budget to Most Likely Costs to achieve more realistic long-term defense acquisition budgets, reduce apparent cost growth and achieve increased program stability.

7. Economic Production Rates should be used whenever possible and advantageous.

8. Assure Appropriate Contract Type in order to balance program needs and cost savings with realistic assessment of contractor and Government risk.

9. Improve System Support and Readiness by establishing objectives for each development program and "designing-in" reliability and readiness capabilities.

10. Reduce the Administrative Cost and Time to Procure Items by raising the limit on purchase order contracts and reducing unnecessary paperwork and review.

11. Incorporate the Use of Budgeted Funds for Technological Risk by quantifying risk and incorporating budgeting techniques to deal with uncertainty.

12. Provide Adequate Front-End Funding for Test Hardware in order to emphasize early reliability testing and to permit concurrent development and operational testing when appropriate.

13. Governmental Legislation Related to Acquisition which unnecessarily burden the acquisition or contracting process should be eliminated.

14. Reduce the Number of DOD Directives by performing a cost-benefit check and requiring that the DAE be the sole issuer of acquisition-related directives.

15. Funding Flexibility should be enhanced by obtaining legislative authority to transfer individual weapon system procurement funds to RDT&E when appropriate.

INITIATIVES ON IMPROVING THE ACQUISITION PROCESS (CONTINUED)

16. Contractor Incentives to Improve Reliability and Support should be developed and introduced into RFPs, specifications and contracts.

17. Decrease DSARC Briefing and Data Requirements in order to increase the efficiency of DSARC and other program reviews.

18. Budgeting Weapons Systems for Inflation should be adopted in order to more realistically portray program cost.

19. Forecasting of Business Base Condition at Major Defense Plants by coordinating interservice overhead data and providing program projections to plant representatives.

20. Improve the Source Selection Process by placing added emphasis on past performance, schedule realism, facilitization plans and cost credibility.

21. Develop and Use Standard Operational and Support Systems to achieve earlier deployment and enhanced supportability with lower risk and cost.

22 Provide More Appropriate Design to Cost Goals to provide effective incentives during early production runs.

23. Assure Implementation of Acquisition Process Decisions by initiating an intensive implementation phase.

24. (ISSUE A) DSARC Decision Milestones should be reduced to "Requirements Validation" and "Program Go-Ahead." See other side.

25. (ISSUE B) MENS should be submitted with Service POM thus linking the acquisition and PPBS process.

26. (ISSUE C) DSARC Membership should be revised to include the appropriate Service Secretary or Service Chief.

27. (ISSUE D) The Defense Acquisition Executive (DAE) should continue to be the USDRE

28. (ISSUE E) The Criterion for DSARC Review should be increased to $200M RDT&E and $1B procurement in FY80 dollars.

29. (ISSUE F) Integration of the DSARC and PPBS Process will be achieved by requiring that fiscally executable programs be presented for DSARC review.

30. (ISSUE G) Logistics and Support Resources will be included in the Service POM by weapon system, and Program Managers will be given more control of support resources, funding and execution.

31. (ISSUE H) Improved Reliability and Support for expedited ("Fast Track") programs will be achieved by requiring an early decision on the additional resources and incentives needed to balance the risks.

32. Increase Competition in acquisition by establishing management programs and setting objectives. (July 27, 1981).

It will undoubtedly be some time before the impact of the initiatives on T&E policy and practice is fully understood. In the meantime, the challenge for top management is to ensure that the new policies are clearly articulated, rapidly disseminated, and effectively implemented. The challenge for individual program managers is to adapt to the current management philosophy while maintaining the stability of existing program plans and acquisition strategies.

EXECUTIVE SUMMARY

One explanation often given for the propensity of the DOD to equivocate milestone decisions is that inadequate test and evaluation data is presented at milestone reviews. This study showed that while most milestone II and III decisions are, in fact, made conditionally, the adequacy of T&E is not the predominant factor. However, when DSARC issues do arise which involve the quantity or quality of T&E data available for decisionmaking, it is generally because of a lack of communication between OSD and the Services. Clarification of milestone review policies, early identification of T&E issues associated with the selected acquisition strategy, and vigorous use of the TEMP are some of the ways this aspect of the acquisition process can be improved in the present environment of decentralized operations and centralized decisionmaking.

The objective of this study was to gain a better understanding of how the adequacy of test and evaluation (T&E) data affects defense system acquisition review council (DSARC) decisions. Specifically, the study focused on the adequacy of the T&E data presented at milestone reviews as a possible reason for conditional decisions. Conditional decisions are those that nominally approve entry into the next phase of acquisition, but also impose restrictions on the pace of development or the rate of production. Such restrictions invariably increase the time and cost of the acquisition process.

Adequacy of T&E data is a difficult concept to define, however, it is important to distinguish between the adequacy of testing and the test results themselves. In this context, adequacy is defined as the quantity and quality of data that is required for an assessment of the system as part of an overall milestone review. Adequacy deals with whether the "right" tests were conducted, under the "right" conditions, in the "right" numbers rather than whether the system "passed" or "failed." (The quotation marks indicate that this is a simplified view of a very complex and frequently subjective determination.)

Numerous attempts have been made to establish precisely the nature and kind of testing required at each stage of the acquisition process. In general, these attempts have not been successful because adequacy is a relative term. Establishing the adequacy of T&E data on a

given program depends on the program's unique circumstances and on the disposition of the defense establishment toward caution or expediency as an operating mode. In addition, as it relates to the defense decisionmaking process, adequacy of T&E data is a strong function of the expectations of the decisionmakers as well as the environment in which the decision is being made. Consequently, for this study, the DSARC principals' own judgment, as indicated by the decision documentation, was taken as evidence of the adequacy of T&E data for each program studied.

Nearly three-quarters of the 38 milestone II and III reviews held between January 1977 and August 1980 resulted in conditional decisions, and limited production, in particular, was a common means of equivocating the milestone III decision. Based on the data collected, it was found that inadequate T&E data did contribute to some conditional decisions, although it was by no means the predominant factor. Breaches of performance thresholds were more common, occurring in 21 of the 28 conditional decisions made during the period under study. In addition, it was found that "hidden agenda" items frequently played a role in the decisionmaking process. Nevertheless, decision documentation on 9 of the conditional decisions reflected a dissatisfaction with the quantity or quality of the T&E data available, and it would be desirable to reduce or eliminate such instances in the future.

Six programs which appeared to involve inadequate T&E data at milestone III were studied in detail. These included the Major Caliber Lightweight Gun, the Aegis Weapon System, the EF-111A Tactical Jamming System, the Air Launched Cruise Missile, and the Patriot and Roland Missile Systems. For each system the milestone review was analyzed in the context of program-specific events as well as the overall decision environment. It was found that several of the programs shared common attributes, including that they were longer than average, that they had attracted a great deal of attention from the GAO and Congress, that in some cases they were still wrestling with fundamental mission need or concept selection issues, and that there was evidence of intraservice disputes over priority or the interpretation of the T&E results. Perhaps most importantly, in 4 of the 6 cases an approved TEMP did not exist early enough to effect the adequacy of T&E data presented to the DSARC.

In the cases studied, inadequate T&E data generally resulted from a lack of communication between OSD and the Services. Either the expectations of the principals were never established, or there was a change in the program or the decision environment which was not communicated to the

other party. Specific underlying causes of inadequate T&E data most frequently included:

- Changes in acquisition policy since the last milestone review which altered the ground rules for proceding from one phase of the acquisition process to the next.

- Shifts in emphasis on selected aspects of the acquisition process, such as a preoccupation with reliability and maintainability, by key players or groups.

- Ambiguities regarding the scope of the system to be tested, including how much of it would be available for OT&E, how representative it would be of the production configuration and how it would be evaluated in an operational environment.

- R&D problems which reduced the time and/or funds available for T&E.

- Problems in conducting the tests themselves due to difficulty in obtaining critical test resources or to inherent range/instrumentation limitations.

Even when there is evidence of inadequate T&E data, other factors tend to dominate the decision process. Particularly at milestone III, there are enormous pressures to proceed with some form of production regardless of the risks or the maturity of the system. These pressures include the national priority of the program, the advancing threat, the long lead times, and the influence of an established constituency with specific vested interests in the program's continuation. On the other hand, there is circumstantial evidence that the adequacy of T&E data sometimes is used as a convenient substitute for other issues including near term affordability problems. In some cases, it may be easier for defense decisionmakers to stretch a program by requiring additional testing, thus spreading the costs over time, then it is to cancel any given program outright.

Another conclusion of the study was that formal communication between OSD and the Services regarding the quantity and quality of T&E required at each milestone decision is critically important in an environment which employs decentralized operations but retains centralized decisionmaking. The Test and Evaluation Master Plan (TEMP) has existed since 1975 but was not rigorously used until about 1979. Although the TEMP is now viewed by some as a "contract" between the parties, its real value

appears to be as a mechanism for insuring continuing structured communication regarding the T&E program. However, it is important that T&E plans as shown in the TEMP accurately reflect the program's response to any changes in the fundamental assumptions to the acquistion strategy.

It was also found as part of this study that improvements are needed in the manner in which T&E data is integrated into the milestone review process. First, the independent T&E assessment by the DDT&E is an important and valuable input to DSARC decisionmaking; however, it is but one part of an overall evaluation that is not presently accomplished in a rigorous fashion. Such an overall evaluation must view the existing T&E data in the context of a formal risk assessment of the alternative courses of action. Second, it was found that many of the questions regarding adequacy of T&E data at milestone III reviews arise out of the immaturity of the hardware or software used for DT&E and IOT&E. Although present policy authorizes initial production for IOT&E along with full scale development at milestone II, there appears to be a number of misconceptions and barriers which have prevented full implementation of this approach. Third, the study showed that a disproportionate amount of management attention is now focused on milestone III, by which time there are few realistic alternatives to proceeding with some form of production. Although there are explanations for this phenomenon, it was found that a better balance between available information and the ability of the bureaucracy to influence program direction occurs earlier in the program about the time of milestone II. In this case, the emphasis of the milestone review necessarily must be on the identification of critical test issues and on the proposed T&E plans rather than on extensive T&E results. Although the study has shown that lack of a common understanding of T&E issues and plans is a fundamental cause of subsequent problems, such a shift in emphasis raises questions regarding the role IOT&E in the decisionmaking process.

Finally, it was found that there are institutional pressures which pose obstacles to the developer and tester in conducting an adequate T&E program. Competition for resources and other interservice-related factors tend to cause success oriented planning without adequate flexibility to deal with inevitable R&D problems. As milestone review dates approach, developers are forced to limit testing in order to solve other problems. Frequently, premature production reviews were held because of a perception that milestone III approval was necessary in order to preserve follow-on production funding. Other institutional idiosyncracies include the fact that changes in program funding due to perturbations in the PPBS process are seldom coupled to official changes to the acquisition

strategy and that the allocation of test range costs under the uniform funding policy sometimes acts as a disincentive to conduct comprehensive, thorough testing.

As shown in Exhibit VI-1, 11 recommendations were proposed in response to the findings and conclusions of the study. These are listed below and discussed in detail in Chapter VI.

- The DSARC should place greater emphasis on earlier milestone reviews in order to influence and optimize the acquisition strategy without major economic or schedule disruptions.

- OSD should clarify the policy regarding the meaning of milestones II & III so that initial production hardware is routinely evaluated during IOT&E.

- New resource programming techniques should be devised which foster the transition from R&D to production and prevent premature production reviews.

- OSD should obtain better consensus regarding the meaning of "Operational Environment" for each major system so that future confusion is avoided.

- More rigorous T&E planning, including the use of system engineering discipline, should be used to define the scope of the system to be tested and avoid ambiguities which lead to differing expectations by defense decisionmakers.

- OSD should reiterate and emphasize use of the TEMP as the principal means of communicating the T&E requirements for proceeding from one phase of acquisition to the next.

- Better overall evaluations should be conducted as part of the milestone review process including more rigorous analysis of alternative courses of action.

- Program Managers should submit impact statements on their programs following changes in acquisition policy so that compliance issues can be surfaced as early as possible. Also, OSD should be more sensitive to documenting changes in emphasis on the acquisition process as it occurs.

- The DDT&E should submit recommendations regarding institutional improvements in T&E after each DSARC in which T&E problems are an issue.

- OSD should continue to monitor the relationship between T&E data and DSARC decisions in order to determine whether present policies regarding the TEMP are effective.

- Research should be undertaken to determine the underlying causes of performance breaches as the predominant contribution to conditional decisions.

TABLE OF CONTENTS

L I S T O F E X H I B I T S

I. INTRODUCTION

I. INTRODUCTION

I. INTRODUCTION

This report describes a research study conducted by Booz, Allen & Hamilton Inc., for the Defense Systems Management College, Ft. Belvoir, Virginia. The objective of the study was to gain a better understanding of how the adequacy of test and evaluation (T&E) data affects Defense System Acquisition Review Council (DSARC) decisions. The study addressed the 38 milestone II and III reviews held between January 1977 and August 1980.

The role of T&E in the decisionmaking process has changed profoundly over the last several decades, and even today is not universally understood. The Department of Defense (DOD) has tried repeatedly to find ways to shorten the acquisition cycle without incurring unacceptable risks. But the risks are elusive without adequate testing and the adequacy of T&E data is a difficult concept to define. However, in defining the concept, it can be stated that "adequate" testing and data is:

- A strong function of the expectations of the decisionmakers
- Related to whether testing is comprehensive, rigorous, and complete
- Different from the assessment of performance
- Only one part of the overall evaluation process.

The DSARC principals may qualify their decisions and thus restrict the pace of development and production because of inadequate T&E data. It is also possible, however, that the adequacy of T&E data is a convenient substitute for other issues that are more difficult to deal with.

This chapter gives a brief background of the problem under study, describes the relationship of T&E data to the DSARC process, enumerates study objectives, and reviews previous analyses.

1. BACKGROUND

In the decade since former Secretary of Defense Melvin R. Laird and Deputy Secretary of Defense David Packard began major reforms and improvements in the

defense systems acquisition process, there has been a litany of proposed approaches to solving the cost, schedule, and performance problems which plagued new weapon systems then as they do today. Among the management theories and styles, the procedures, and documentation, two themes emerge. The first is that a sort of dynamic tension has always existed between those managers who would accelerate system developments at all costs (and, knowingly or not, accept the risks therein) and those who would proceed in a more cautious, step-by-step fashion. The latter group, while certainly not opposed to accelerating the acquisition process, and not unmindful of the costs associated with long development times, simply takes a broader view of the ultimate cost of acquiring a weapon system, i.e., the cost to do it right versus the cost to do it over.

Not everyone agrees that these positions are mutually exclusive; however, most observers of the acquisition process agree on the second theme: that the mood of the defense establishment invariably swings like a pendulum toward one extreme and then the other. During the 1960s, for example, concurrency of development and production was an accepted technique for reducing development time. During the 1970s, it was officially proscribed due to numerous technical failures which were experienced. Now, with the March 1980 publication of DOD Directive 5000.1 and DOD Instruction 5000.2, the technique is cautiously being reintroduced. Other examples of shifts in management philosophies include centralized and decentralized decision-making, total package procurement and "fly-before-buy," government direction of contractor efforts, and hands-off strategies, etc. At one point, the initial operational capability (IOC) was considered a parameter subject to tradeoffs just like performance and cost. Now, accelerated deployment of systems seems to be the dominant theme.

T&E is the pivot on which this pendulum swings. Some see T&E as the only means of objectively measuring the ability of a system to meet an expressed need and as the primary vehicle for deciding when to make progressively larger investments in response to diminishing risks. Others see T&E principally as feedback to the design process and, except in extreme cases, largely unrelated to the decision to proceed to the next phase of acquisition. This view of T&E is caused by a perception that it generally would be too costly and time consuming to do otherwise. The pressure on the one hand is to go faster and, on the other hand, to avoid making a mistake.

The DSARC is the DOD body charged with advising the Secretary of Defense, on a case-by-case basis, which of these two paths is most desirable. While historically the number of outright cancellations recommended by the DSARC has been small, there has been a trend in recent years to equivocate the decision to proceed, i.e., to stretch out programs and proceed at a slower pace while problems are resolved, or to attach qualifications, often involving additional test and evaluation, to the systems' continued development or production. Some view such "T&E strings" simply as micromanagement by OSD; others see them as substitutes for more difficult issues. In either case, this situation invariably lengthens the acquisition cycle.

In the 38 milestone II and III decisions made between January 1977 and August 1980, nearly three-quarters imposed conditions of one sort or another. The school of thought that would favor faster development would take strong exception to the additional development costs, uneconomical procurement buys, and the stretched out IOCs that resulted. However, the school favoring demonstrated performance would be frustrated by the frequency that the Services brought unsatisfactory items to the DSARC for review. Both schools can only be dismayed by the inefficiency of the process and hope that it can be improved.

2. T&E DATA AND THE DSARC

The defense establishment instituted a formal milestone review process more than 10 years ago and has continuously refined it. Yet, DSARC reviews continue to reveal deviations from plans, performance failures, affordability issues, and other unexpected occurrences. One explanation might be that if everything passed DSARC review, then DOD simply was not being as aggressive or ambitious as it should be in its selection of projects. After all, some failures must be expected if doctrinal and technical barriers are to be overcome. However, a more realistic view in light of the severe budget constraints facing DOD is that there are too many deviations from agreed-to plans and that "disconnects" are occurring between the Services who are developing the systems, and the OSD who is reviewing them.

DSARC decisionmaking must be viewed as an amorphous process which is affected as much by the predisposition and personalities of the principals, and by exogenous events and the tenor of the times, as it is by the isolated facts regarding the sytem under review. Nevertheless, since the later 1970s, there has been increasing emphasis on supporting milestone decisions with hard data

and on minimizing the need for subjective judgements on both developmental and operational matters. The responsibility for generating these data falls on both the project office and the test agencies.

One theory that is often expressed as an explanation for the propensity of the DSARC to qualify its decisions is that inadequate testing is done or that inadequate T&E data are available in support of the milestone review. The premise is that a milestone review involving inadequate test data frequently results in a conditional decision which imposes requirements for additional testing, presumably to make the body of test data more adequate.

3. STUDY OBJECTIVES

This study focused on the issue of adequacy of T&E data as it relates to the decisionmaking process. In particular, the study attempted to determine whether inadequate T&E data result in conditional decisions and, if so, what were the underlying causes of inadequate data. Research conducted during the course of the study also explored other possibilities for explaining the conditional DSARC decisions and the way adequacy of testing is treated in the milestone review process.

The following specific questions guided the research efforts:

- Are conditional decisions caused by inadequate T&E data? Or are there other explanations?
- If conditional decisions are caused by inadequate data, what are the underlying causes of such situations?
- Is there effective communication between OSD and the Services with respect to the quantity and quality of T&E data required for milestone decisions?
- Is the adequcy of T&E data effectively addressed in the present milestone review process?
- What are the significant obstacles facing the developer and tester in planning and conducting an adequate T&E program?

4. PREVIOUS ANALYSIS

In 1972, the General Accounting Office (GAO) issued its first major report on T&E problems in the DOD. Entitled The Importance of T&E in the Acquisition Process

for Major Weapon Systems, the report concluded that most systems studied did not have adequate test plans and that T&E was not accomplished in a timely manner. Most importantly, the GAO also concluded that T&E data were not available prior to those times in the acquisition cycle when decisions had to be made. This report may have contributed to the "T&E explosion" which occurred during the 1970s. However, the reform in T&E policy and the increased emphasis on T&E as a means of improving the acquisition process actually began with the President's Blue Ribbon Defense Panel in July 1970 and the Deputy Secretary of Defense memos on T&E to the Service Secretaries between February and August 1971.

The Blue Ribbon Defense Panel was one of the first groups to articulate problems with Total Package Procurement policies that were then in effect. The panel stressed the importance of periodic review of acquisition programs at critical milestones. These reviews would provide an opportunity to decide to continue, reorient, or cancel a program based on a comparison of that program's progress to previously established goals and objectives. The impact on test and evaluation was obvious. In order to be effective, decisionmakers needed reliable and accurate measurements of program progress compared to program objectives.

The coming emphasis on T&E was anticipated somewhat by Deputy Secretary Packard. In May 1969, he established the DSARC as the formal DOD acquisition program review body and, in mid-1971, he issued a series of memoranda which established a focal point for T&E within the Deputy Director, Research and Engineering (DDR&E). He also required each Service to establish independent test agencies and strong headquarters staff focal points, and generally elevated the importance of T&E in the acquisition process.

Since then, there has been continued interest in and revision of DOD T&E policy. Numerous groups and individuals, both internal and external to Government, have searched for solutions to what one researcher called "The Persistent Problem." The Defense Science Board (DSB) established a Task Force on Test and Evaluation in 1974 which developed guidance on test and evaluation through examination of general issues and the generation of checklists oriented toward good procedures and practices relative to T&E. At the same time, the Deputy Director (Test and Evaluation) published a series of reports establishing T&E guidelines for various types of weapon systems. In 1977, the DSB formed a Task Force on Test and Evaluation Policies which explored broader questions

regarding the role of T&E in the acquisition process. However, these and other studies have not focused on inadequate T&E data as a possible explanation for conditional program decisions. A selected annotated bibliography of related studies is contained in Appendix A.

II. STUDY APPROACH

II. STUDY APPROACH

DSARC decisions represent a unique blend of technical, economic, and political factors. Consequently, the technical approach used for this study was to categorize and define the apparent reasons for conditional decisions and the underlying causes of inadequate T&E data based on program unique issues and the overall environment in defense program decisionmaking for the period of interest. As shown in Exhibit II-1, the study progressed in five steps:

- Define the study objectives
- Identify and characterize DSARC decisions for the period January 1977-August 1980
- Describe the decision environment with respect to policy and external events
- Select the subset of decisions which seem to involve inadequate T&E data for in-depth study in the context of the decision environment; identify common elements, underlying causes and other possible reasons for the decision and its consequences
- Integrate the findings and draw conclusions and make recommendations for improving the situation in the future.

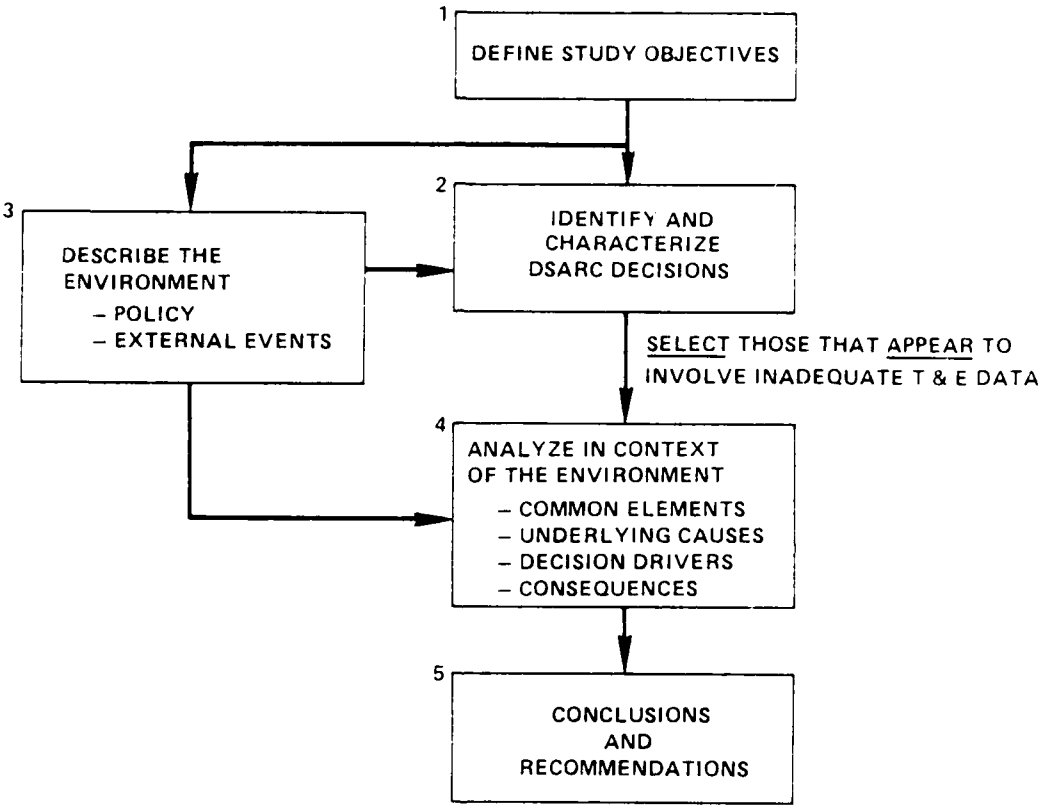
This chapter presents some definitions and establishes a framework for discussing the DSARC decision process. Data collection and analysis techniques that were used are also described.

1. DEFINITIONS

The primary goals of DOD T&E activities are the gathering and analysis of test data and the formulation of assessments that support the acquisition decision process.* Identification of critical issues with measurable goals and thresholds and statistically significant sample sizes is a worthy, if sometimes impractical, objective. However, when asked to define adequacy of T&E data, one

* Introduction to the Application of Statistical Concepts to Test and Evaluation. Office of the Director, Defense Test and Evaluation, August 1980.

EXHIBIT II-1
Technical Approach



OSD official commented that the DSARC was looking for "...reasonable evidence of a reasonable amount of testing that was reasonably successful." While imprecise, this definition is probably very nearly an accurate description of what actually happens when the DSARC principals try to integrate the hard and soft data presented to them.

Numerous attempts have been made to define in absolute terms the amount and kind of test and evaluation required during each phase of the acquisition cycle. For example, the DDT&E guidelines for various types of weapon systems provide excellent checklists of things that should be considered. However, financial, temporal and other pressures almost always restrict the amount of testing that is realistically possible and a generally accepted formula has never been devised for conducting trade-offs between testing and other program activities. Thus, in the final analysis, the determination of adequacy of testing in a given case rests with the decision makers themselves who must weigh the costs and risks of various alternatives. Therefore, for this study, the DSARC principals' own statements are taken as evidence of the adequacy of T&E data, i.e., the adequacy of T&E data is defined as:

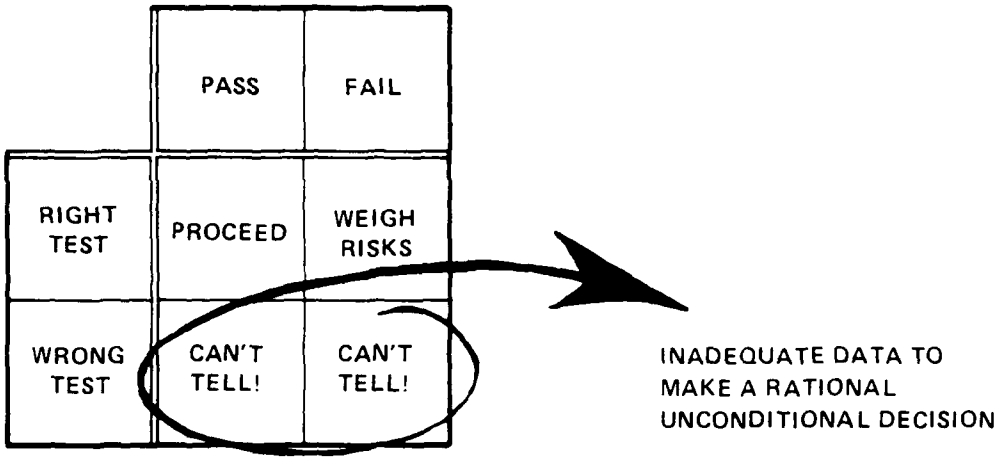
> The quantity and quality of data which in the judgement of the DSARC principals are necessary for an assessment of the system in the context of an overall evaluation.

Note that this definition does not address the inherent information content of the data. That is, the adequacy of the testing is differentiated from the test results themselves. Th s point is shown graphically in Exhibit II-2.

Viewed simplistically, tests can be divided into two classes, right and wrong, where wrong tests include those that are not comprehensive, rigorous, or complete. The outcome of a test can be broadly classified as either pass or fail, depending on the evaluation criteria, the inherent quality of the system design, and the interpretation of the results. This leads to a two-by-two matrix of possible events:

- The right tests are conducted and the system passes, in which case, all other things being equal, the decision is rather straightforward and an unconditional decisions is made to proceed to the next phase.

EXHIBIT II-2
Adequacy of Testing vs. Test Results



"WRONG" MEANS THOSE TESTS THAT ARE NOT COMPREHENSIVE, RIGOROUS OR COMPLETE

- The right tests are conducted, but the system fails all or some portion of them. This is the arena of management judgement in which the risks of proceeding, in the hopes of making the system work, are weighted against the risks associated with delaying, fixing, and retesting. This is the classic case in which defense leaders routinely disagree because of differences in their own perspectives and attitudes.

- The system apparently passes, but the tests are judged by the DSARC to be the wrong ones. Here the situation is not so clear, even with regard to the structure for debating alternative courses of action. While it is possible to define the risks in not proceeding (increased cost, advancing threat, etc.), depending on how wrong the tests are, there are inadequate data to rationally assess the risks in proceeding. For example, it might be difficult to assess the risks in proceeding with a system that is required to operate in an electronic countermeasures (ECM) environment if it had not been tested under ECM conditions.

- The last case, where the system fails the wrong tests, is nearly identical to the preceding. Depending on whether one was an advocate or an antagonist, it would be easy to be critical of a system with bad test scores. Unfortunately, if the tests were somehow inappropriate or improper, this might be overly harsh and result in holding a system back unnecessarily. For example, testing the ability of a single aircraft to penetrate enemy air defenses when the intended tactic is to penetrate in mass or in conjunction with other systems might make difficult a rational decision to proceed. [It must be pointed out, however, that these data might be very significant in an overall evaluation because it would bound the expected performance of the system.]

The last two cases involve inadequate T&E data; i.e., the tests themselves are determined by the DSARC principals to be somehow lacking and the resulting data are judged to be inadequate to make an unconditional decision to proceed.

The adequacy of T&E data involves both quantity and quality. Quantity deals with the probability of success, confidence limits, and the incremental worth of collecting additional data versus the risk of making an incorrect decision. Quality issues deal with the realism of the test (e.g., testing Copperhead under realistic battlefield conditions such as smoke and fog) or the lack of specific technical or operational tests thought to be important (e.g., testing the F-16 at the extremes of its performance envelope or testing Patriot's compatibility with related air defense and command and control equipment).

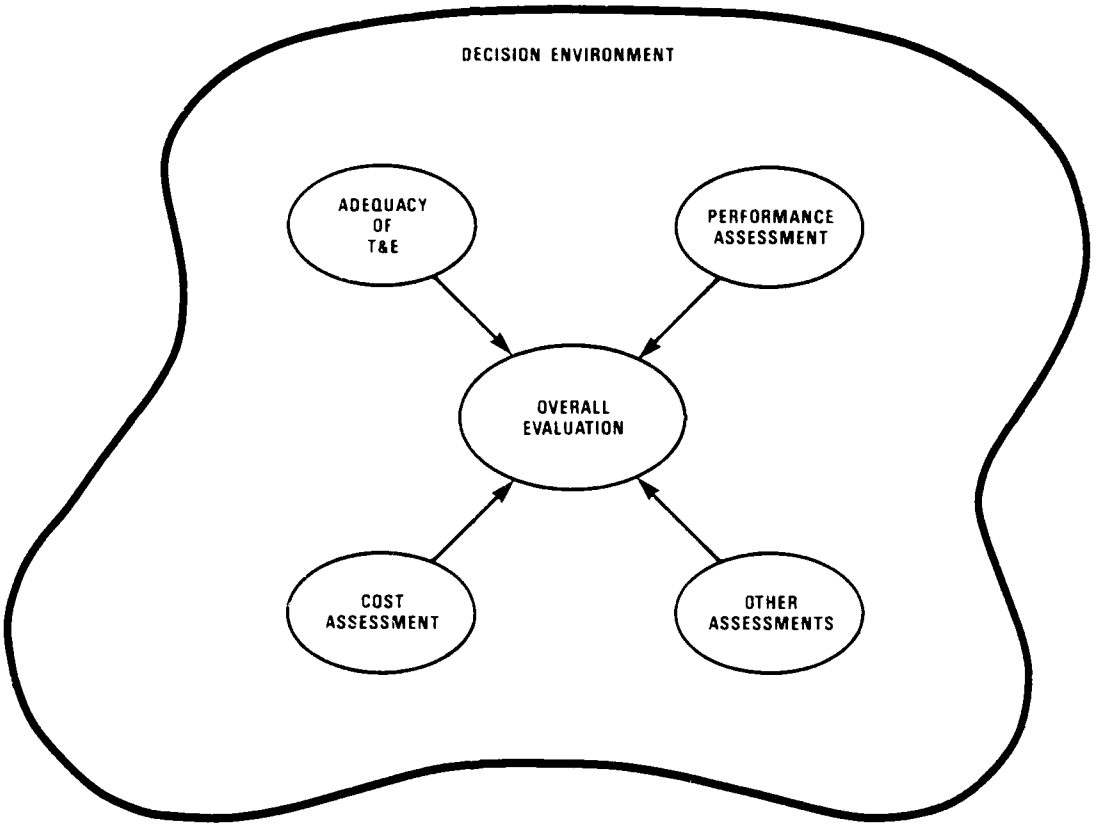
2. FRAMEWORK FOR DSARC DECISIONS

Exhibit II-3 shows a model of the DSARC decision process in which various types of assessments and the overall evaluation take place in and react with the decision environment. The decision environment includes the established policies, the attitudes of defense decisionmakers, and other external events, such as Congressional involvement or decisions on other related programs. Any of the assessments can lead to conditional decisions. For example, a judgement that there is inadequate T&E may result in delays in order to conduct additional tests; a performance assessment which indicates a failure to meet a particular threshold may result in a fix and retest cycle; and other assessments may result in slowing the pace of the program for other specific reasons.

Within this framework a possible syntax for discussing DSARC decisions is shown below:

The syntax proposed is that various decisions are caused by one or more reasons which, in turn, are due to underlying causes. The decision itself can be of several types, such as unconditional approval to proceed to the next phase, varying degrees of conditional approval, or denial or delay in proceeding including program cancellation. For each decision type, there are a number of apparent reasons generally given by the DSARC principals as justification.

EXHIBIT II-3
DSARC Decision Model



The set of apparent reasons can be defined by looking at both the management and system issues that are reviewed at milestones II and III (Exhibit II-4). It can be seen that the issues group into four general areas: need, affordability,* maturity, and management/planning. Failure to address each of these issues satisfactorily can constitute a reason for less than unconditional approval.

These issues appear logical in terms of a rational dialogue concerning risk and investment of additional resources. In order of importance, they address the following questions:

- Does the system address a legitimate need?
- Is the system affordable?
- Is it sufficiently mature and will it likely meet its technical and operational goals?
- Are the program management and planning aspects of the program well in hand?

While T&E touches on all four categories, it is most frequently raised as an issue relating to maturity and, less often, to management and planning.

If the apparent reasons for a particular decision have been determined, then an attempt can be made to identify and understand the underlying causes. For example, one possible underlying cause of a failure to present adequate T&E data to the DSARC (a maturity failure) is poor preparation or coordination of the test and evaluation master plan (TEMP). Another is the lack (or availability) of appropriate test facilities or instrumentation. A third might be insufficient time or money to conduct a comprehensive test program.

---

* There are many definitions of affordability; however, the one adopted here is simply the ability to provide adequate resources to acquire and operate a system in an efficient and effective manner. The "bow wave" (discussed later) refers to the financial condition in which more systems are developed than can be procured and deployed within anticipated defense budgets.

	DSARC II	DSARC III
NEED	• MISSION ELEMENT NEED/TASK REAFFIRMED AND INTERACTIVE THREAT ANALYSIS UPDATED • SYSTEM SELECTED MEETS NEED AND IS COST-EFFECTIVE	• MISSION ELEMENT NEED REAFFIRMED AND INTERACTIVE THREAT ANALYSIS UPDATED
AFFORDABILITY	• AFFORDABILITY CONFIRMED; LIFE-CYCLE COSTS WITHIN AMOUNTS REFLECTED IN LATEST FYDP/EPA	• SCHEDULE AND COST ESTIMATES REALISTIC • LIFE-CYCLE COSTS WITHIN AMOUNTS REFLECTED IN FYDP/EPA • PROGRAM AND FISCAL YEAR THRESHOLDS REAFFIRMED
MATURITY	• TRADE-OFFS BETWEEN COST, PERFORMANCE, SCHEDULE AND LOGISTIC SUPPORTABILITY ARE ACCEPTABLE • UNCERTAINTIES/RISKS ARE IDENTIFIED AND ACCEPTABLE • SYSTEM REQUIREMENTS ADEQUATELY SPECIFIED AND INCLUDE PERFORMANCE DESIGN-TO-COST AND LIFE-CYCLE COST NUCLEAR SURVIVABILITY ELECTROMAGNETIC COMPATIBILITY ELECTRONIC (INFRARED, OPTICAL) COUNTERMEASURES PRODUCIBILITY – NATO STANDARDIZATION AND INTEROPERABILITY LOGISTICS SUPPORTABILITY • RELIABILITY AND MAINTAINABILITY GOALS AND THRESHOLDS ESTABLISHED • MANPOWER GOALS AND THRESHOLDS IDENTIFIED	• DEVELOPMENT PROGRESS AND OT&E RESULTS SUPPORT DECISION TO PROCEED COST, SCHEDULE, PERFORMANCE, AND SUPPORTABILITY TRADE-OFFS ACCEPTABLE DESIGN TO COST AND LIFE-CYCLE COST REQUIREMENTS REALISTIC SYSTEM COST-EFFECTIVE AND AFFORDABLE; REMAINS BEST ALTERNATIVE – MAJOR PROBLEMS IDENTIFIED/RESOLVED – NATO STANDARDIZATION/INTEROPERABILITY REQUIREMENTS SATISFIED • SUPPORT SUBSYSTEMS FOR INITIAL OPERATIONAL UNITS ASSESSED • MINIMUM OPERATIONAL R&M MET; GOALS AND THRESHOLDS ESTABLISHED FOR INITIAL DEPLOYMENT • MANPOWER AND TRAINING REQUIREMENTS DEVELOPED
MANAGEMENT ISSUES AND PLANS	• UPDATED ACQUISITION STRATEGY SUPPORTS PROGRAM OBJECTIVES AND IS BEING PROPERLY EXECUTED • PROGRAM MANAGEMENT STRUCTURE SOUND AND ADEQUATELY SUPPORTED • PROBLEMS MANAGEMENT PLANS COMPLETE AND INCLUDE – APPROACH TO RESOLVE REMAINING UNCERTAINTIES AND RISKS – REALISTIC FALL BACK ACTIONS AND ALTERNATIVES – BUSINESS PLANS TO SUPPORT STRATEGY; CONTRACT TYPES CONSISTENT WITH OBJECTIVES – COMPETITIVE SELECTION OF SUBSYSTEMS; CONSIDERATION OF EXISTING MILITARY OR COMMERCIAL HARDWARE/SOFTWARE – CONSIDERATION OF FOREIGN DEVELOPMENTS – REQUIREMENTS FOR LONG-LEAD PROCUREMENT AND INITIAL LIMITED PRODUCTION; COORDINATION (WHEN APPROPRIATE) WITH SINGLE MANAGER FOR CONVENTIONAL AMMUNITION • FIRM AND REALISTIC COST, PERFORMANCE AND SCHEDULE ESTIMATES, AND THRESHOLDS ESTABLISHED • READINESS TO SUBMIT INITIAL SAR	• ACQUISITION STRATEGY UPDATED AND BEING EXECUTED • PROGRAM MANAGEMENT STRUCTURE AND PLAN SOUND AND ADEQUATELY SUPPORTED • PRODUCTION QUANTITY REQUIREMENTS VALID • PRODUCTION, PRODUCIBILITY, QUALITY ASSURANCE, AND FACILITIES ISSUES IDENTIFIED/MANAGED PRODUCTION READINESS REVIEW COMPLETE – REQUISITES FOR FUTURE DECISIONS DEFINED COMPETITION CONSIDERED; E.G., SECOND SOURCE • PLANNING FOR DEPLOYMENT ADEQUATE MANPOWER AND TRAINING LOGISTICS READINESS INTEGRATION WITH EXISTING SYSTEMS

However, other less obvious reasons also might be involved. For example, overall resource constraints in the planning, programming, and budgeting system (PPBS) program objectives memorandum (POM) could in fact determine the pace at which the Services can place systems that are otherwise ready in a full-scale development or production status. This is the heart of the so-called affordability problem in which there are simply not enough resources to fund all of the Services' proposed programs. Delaying or juggling the pace at which systems are permitted to evolve may be accomplished under the guise of "additional test and evaluation needed" rather than dealing with the harder decision of major cancellations and wholesale deletions. Additionally, technical failures and ensuing barrages of criticism about one program might have a ripple effect on many other programs, far out of proportion to technical needs or the adequacy of T&E data available for review.

Thus, it is important to understand the overall environment as well as the program-specific issues at each milestone review. This constitutes the essence of the approach used in this study: to categorize and define the apparent reasons for conditional decisions and to identify the underlying causes of inadequate T&E data based on program-unique issues against a descriptive backdrop of the overall environment in defense decisionmaking.

3. DATA COLLECTION

Data collection consisted of examining the official documentation surrounding each milestone review and structured interviews with Government personnel familiar with specific programs as well as those knowledgeable about T&E in general and the defense decisionmaking process.

Official documents reviewed included the decision coordinating paper (DCP, formerly the development concept paper), the TEMP, the T&E assessment prepared by the Deputy Director, Test and Evaluation (DDT&E), and the Secretary of Defense or Deputy Secretary of Defense decision memorandum (now known as the SDDM).

The DCP has been viewed historically as the contract between OSD and the Service regarding the goals and objectives of the program, the acquisition strategy, and the thresholds and constraints that must be met before proceeding to the next acquisition phase. More recently, it has been used by the DSARC as the primary documentation in arriving at milestone recommendations. Conversely, the TEMP previously was a coordinating document but now serves

in a contract-like capacity similar to the old DCP. Theoretically, the TEMP requires prior agreement on the nature and scope of the tests to be conducted and establishes for all parties the T&E data requirements for each succeeding milestone review. The T&E assessment is an independent, after-the-fact input to the DSARC and the Secretary of Defense regarding the T&E accomplished and planned. The SDDM, among other things, provides guidance and direction for the next phase of acquisition.

Interviews were held with over 35 individuals on the OSD staff, at the Service Headquarters, in project offices, and in test and evaluation organizations. In some cases, a formal interview, using a structured interview guide, was conducted; in other cases, informal discussions were held. Appendix B contains a copy of the interview guide that was used for formal interviews.

4. ANALYSIS

In order to place each milestone decision in its proper historical perspective, trends in acquisition and T&E policy were documented and a diachronic chart covering the decade of the 1970s was prepared. This chart, which is shown in Appendix D and described in detail in Chapter III, identifies external events that could have affected individual decisions. The process of relating such events to their historical context and to each other has proven to be a valuable research technique on similar studies.

The DSARC decisions were treated first as a group in which trends were identified and relationships with the decision environment were analyzed. Then, selected DSARC decisions were analyzed in depth to discover alternative reasons for conditional decisions and the underlying causes of inadequate T&E data.

* * * * *

The fundamental approach used in this study was to analyze those decisions in which the DSARC principals indicated there were inadequate T&E data in the context of the existing decision environment. The next chapter describes that environment in terms of acquisition policy, T&E policy and external events.

III. THE DECISION ENVIRONMENT

III. THE DECISION ENVIRONMENT

The decision environment affects milestone reviews in subtle but important ways. Policy emphasis swings between extremes of caution and expediency and can sometimes find a program out of phase with the dominant philosophy at the time of the review. High-level personnel change as part of the political process and new ideas and approaches become vogue and then are replaced by others. DSARC decisionmakers are influenced not only by external pressures from Congress, the GAO, and the public media, but also by specific events such as the reliability failures of the U.S. helicopters in Iran or the cancellation of the B-1 bomber. Though the environment is complex and highly coupled, it nevertheless provides clues to understanding the DSARC decisions.

This chapter traces the major themes in acquisition and T&E policy since about 1970. In addition, a diachronic chart is used to provide a historical perspective of significant events and trends.

1. TRENDS IN ACQUISITION POLICY

The T&E revolution which occurred in the early 1970s was a by-product of the acquisition policy initiatives of Deputy Secretary of Defense Packard. Appendix C contains a brief overview of some of the key studies and other events since 1970 that influenced acquisition policy. Several important trends can be identified which affect the questions raised in this study regarding adequacy of T&E data. Specifically, these include:

- Centralization of operations within OSD reached a maximum in the mid-1960s, but has been gradually decreasing since the early 1970s and further decentralization can be expected.

- The acquisition reforms embodied in OMB Circular A-109 began in the early 1970s, may have peaked in the late 1970s, and now are leveling off.

- The role of T&E in the decisionmaking process grew rapidly in the early 1970s as a companion to the new way of doing business introduced by Packard. However, as indicated by the Defense Science Board, the

potential exists for testing to increase the length of the acquisition cycle, and the influence of T&E in the decision process may also level off as the desire to expedite development increases.

- The emphasis on operational suitability has risen steadily throughout the last decade. Fueled by increases in weapon system complexity, the all-volunteer force, decreasing skills of enlisted personnel, and revelations about U.S. military readiness, it will probably continue to do so throughout the 1980s.

- Finally, the total rejection of concurrent development and production existing since 1970 has given way to a more permissive attitude in which decreases in acquisition time are traded for increases in risk on a case-by-case basis.

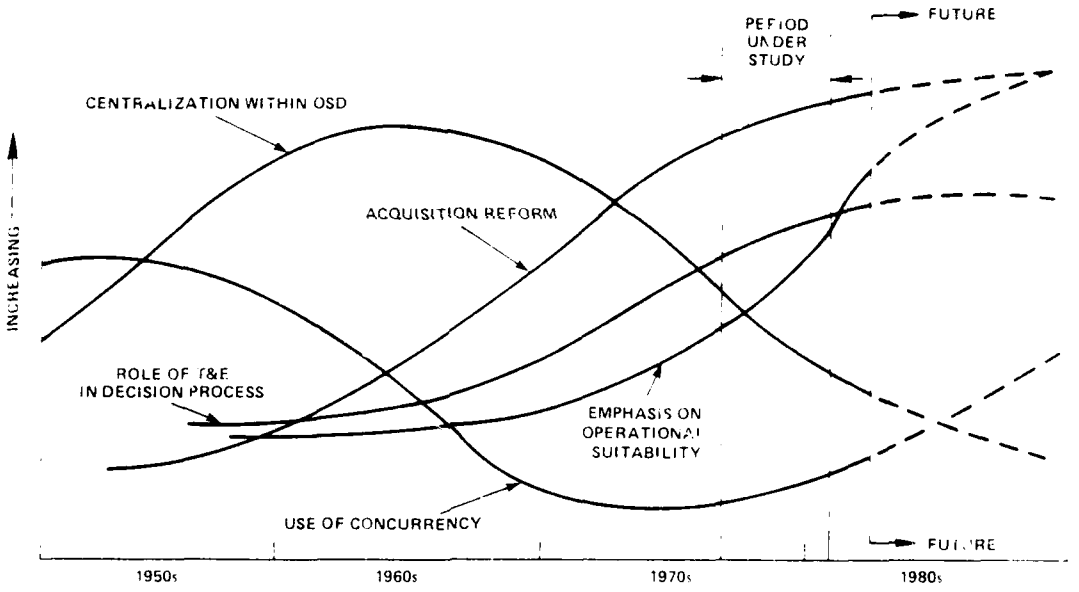
Exhibit III-1 shows these trends on a subjective scale increasing on the y-axis. They provide background for a discussion of the evolution of T&E policy.

2. EVOLUTION OF T&E POLICY

Prior to the 1960s, well-accepted and understood institutional mechanisms of long standing within the Services allowed the agencies concerned in the acquisition process to work out the engineering development and associated testing activities beyond the purview of higher authority. When a product was ready, the appropriate Service testing agency formally evaluated the system for operational use. Deficiencies (if any) were then corrected and the item was approved for Service use by the Service Chief or Service Secretary. This was followed by quantity production.

As described in the previous section and in Appendix C, during the decade of the 1960s sweeping changes to the acquisition process were accompanied by extraordinary centralization of authority and decision-making. The growth of the OSD and Service headquarters staffs in size and involvement in materiel matters was accompanied by an erosion of the cadre of experienced technical personnel and authority of the Services' own materiel development establishments. In former years, these establishments managed most of the test and evaluation materiel acquisition activities as they saw fit, often beyond the effective control of the Service

EXHIBIT III-1
Trends in Acquisition Policy



Chiefs or even the Secretaries. During the early 1970s and continuing to the present, the trend has been reversed. The policy now is to ensure greater decentralization of material acquisition authority with the Secretary of Defense retaining only four key milestone decisions for major programs.

Exhibit III-2 shows the changes in T&E policy that accompanied the larger movement in acquisition policy overall. The time scale is divided according to the introduction and several revisions of DOD Directive 5000.3, i.e., pre-1970, 1970 to mid-1975, mid-1975 to mid-1978, mid-1978 to 1979, and 1979 to the present. The date of the operant DOD directive for each period generally divides the time span in recognition of the fact that management philosophies tend to emerge and even be put into practice somewhat before they are formally introduced as policy.

(1) Baseline

Prior to 1970, the role of T&E in the acquisition process was stated as part of DOD Directive 3200.9, Initiation of Engineering and Operational System Development, i.e., there was no separate DOD directive on T&E, per se. Emphasis at the time was on total Package Procurement, and there was no need to focus on T&E from a decisionmaking standpoint because formal periodic reviews of major systems and the use of demonstrated performance as the pacing function had not yet been introduced. Responsibility for T&E was generally assigned under the DDR&E and received a predominantly developmental emphasis. Although test planning was included as part of the Technical Development Plan (TDP) prepared by the Service, the OSD staff frequently were involved in the details of program management.

Development T&E was judged by the Blue Ribbon Defense Panel as being generally well understood and executed as part of the system design process. Operational T&E, on the other hand, received little emphasis before production; it was judged to be infrequent, poorly designed and executed, and generally inadequate. In addition, the importance of realistic test environments that closely approximated the actual operating conditions of the system was not fully appreciated.

EVOLUTION OF TEST &

TIME FRAME	MAJOR DOD DIRECTIVE	BASIC PRINCIPLES	OSD RESPONSIBILITIES	TEST PLANNING	
Pre 1970	Overall acquisition policy stated in DODD 3200.9 No local DOD directives on T&E per se	Emphasis on "total package" procurement with little use of demonstrated performance as the pacing function; no formal periodic reviews	T&E responsibilities generally assigned under the DDR&E; received a predominantly developmental focus	Test planning included as a chapter in the Technical Development Plan (TDP); OSD frequently involved in detailed direction of programs	Assess... generall...
1970-Mid 1975	DODD 5000.3 January 9, 1973[1] (8 pages)	Disciplined use of T&E results to assist in reducing risks • Conduct T&E early • Base schedules on accomplishing T&E prior to key milestone decisions • Define need and estimate military utility prior to initiating development	DDT&E established and designated as having across the board responsibility for OSD in T&E matters. Provides to SECDEF and DSARC independent assessment of T&E accomplished and planned	The DOD component prepares an overall integrated T&E plan as early as possible (no OSD approval required). The plan is to be kept current	Conducted... risks ha... specifica... • Planned... • Start ... prototy... • Specifi...
Mid 1975-Mid 1978	DODD 5000.3 May 20, 1975[2] (8 pages)	SAME AS ABOVE	SAME AS ABOVE	The test plan is now identified as a separate document called a "Test Master Plan (TEMP)." Required approval is ambiguous. Changes are to be documented	
Mid 1978-1979	DODD 5000.3 April 11, 1978 (16 pages)	Role of T&E elaborated and expanded • Increased emphasis on operational effectiveness and suitability • Issues, test criteria and measures of effectiveness established before tests begin • Allow for contingency testing • Minimize subjective judgment	Staff responsibilities within OSD split • DDT&E for DT&E and coordinates OT&E aspects • USD (PA&E) for OT&E matters	TEMP to be submitted to DDT&E for coordination/review relative to major milestones but an approved TEMP is a prerequisite for Milestone II. Changes are to be properly documented	Essentially... testing, ... checklist o...
1979-Present	DODD 5000.3 December 26, 1979 (18 pages)	Essentially the same • Early DT&E and OT&E to reduce risks and estimate operational potential • Issues, objectives and criteria before test begins relate to mission need • Successful T&E required before committing resources or progressing to next phase • Minimize subjective judgment	DDT&E (again) given total responsibility for T&E • Policy and interservice problems • T&E assessments • Joint testing • Major range and test facilities base	TEMP must be submitted for OSD approval prior to Milestone I – guidelines for preparation included in 5000.3. Changes to be reported promptly	Essentially... will suppo...

NOTES

1 While the policies shown were formally introduced with the first publication of DODD 5000.3, much of their substance was issued earlier by DEPSECDEF multi-addressed memoranda
"Conduct of Operational Test and Evaluation," February 11, 1971
"Role of DDR&E in Test and evaluation as Related to DCP System," April 21, 1971
"Test and Evaluation in the System Acquisition Process," August 3, 1971

2 These changes were formally made as 5000.3 Ch1, April 1974

TEST & EVALUATION POLICY

DG	DEVELOPMENTAL T&E	OPERATIONAL T&E	INDEPENDENT EVALUATION	COMBINED DT AND OT
or in the Technical equently involved in	Assessed by the Blue Ribbon Defense Panel as generally well understood and executed	Little emphasis on operational testing before production OT was infrequent, poorly designed and executed and generally inadequate *Importance of realistic test environments approximating* actual operating conditions not generally recognized	No specific policy regarding independent evaluation developing and or using agencies assess performance of new weapon systems	A frequent practice largely due to urgency to deploy and resource limitations for DT&E Consistent with OSD emphasis on concurrency of development and production
verall integrated T&E pproval required) The	Conducted to demonstrate design process is complete risks have been minimized system will meet specifications • Planned conducted monitored by developing agency • Start early include components subsystems prototypes and preproduction models • Specific objectives for each phase	Importance of operational testing begins to emerge Initial OT&E (IOT&E) to be conducted before major production decision IOT&E to be in as realistic an environment as possible and conducted by personnel like those expected to use and maintain the system Generally, pilot production items to be employed	Each service is to provide one (or a limited number) major field agency responsible for OT&E separate and distinct from developing procuring command Also an OT&E focal point to be established within headquarters staff	OT should be separate from DT however may be combined with early OT when appropriate
separate document P) Required approval is inated	SAME AS ABOVE	Emphasis on earlier OT&E also preproduction prototypes rather than pilot production items are generally to be used for IOT&E	Policy changed such that only one major field agency responsible for OT&E is permitted must be separate from using command as well as developing procuring	SAME AS ABOVE
r coordination review approved TEMP is anges are to be	Essentially the same adds Phase 0 and post Milestone III testing, stresses importance of software and provides checklist of "ilities"	Essentially the same specifically adds enemy countermeasures as part of a realistic environment Also states that preproduction items or full scale engineering models to be used for IOT&E if they are reasonably representative Defines OT&E objectives for each phase	Essentially the same more thorough definition of the role and responsibilities of the independent OT&E agency	Position softened somewhat generally to be conducted separately but may be combined when clearly identified cost schedule benefits or risk acceptable However, normal practice is that OT supporting a production decision *will be conducted independently*
approval prior to ation included in mptly	Essentially the same adds that the development agency will support OT&E objectives as appropriate	Essentially the same adds the objective to obtain a valid estimate of the user's capability to operate and maintain the system in peacetime and wartime	ESSENTIALLY THE SAME	Specifically highlighted that DT and OT may be combined but that the OT&E agency will be a full partner and that OT credibility will be maintained Final testing prior to Milestone III will emphasize operational testing

ulti-addressed memoranda

EXHIBIT III 2
EVOLUTION OF T&E POLICY

Development testing (DT) and operational testing (OT) were frequently combined due to a perceived urgency to deploy new systems and to resource limitations for operational test and evaluation (OT&E). This was consistent with the existing philosophy of concurrent development and production. There were no specific policies regarding independent evaluation of test results, and assessments of new systems were usually performed by the developing or using agencies.

By today's standards, Congress was hardly involved in the acquisition process although they were becoming increasingly concerned with chronic cost overruns and performance deficiencies.

In 1971, Secretary Packard issued three memoranda to the Service Secretaries on T&E which were consistent with his earlier initiatives on overall acquisition policy. These led to the first publication of DOD Directive 5000.3 in January 1973. Since then the directive has been revised three times in response to changing emphasis in acquisition policy and in the perceptions of the role of T&E.

(2) Basic Principles

The first major change in basic principles called for the disciplined use of T&E results to assist in reducing risks associated with moving from one phase of the acquisition cycle to the next. T&E was to be conducted early and development schedules were to be based on accomplishing the required T&E prior to key milestone decisions. These basic principles continue today except that the role of T&E has been further defined and expanded. Emphasis is placed on early development, test, and evaluation (DT&E) and OT&E, particularly to evaluate operational effectiveness and suitability. Issues, objectives, and criteria relating to mission needs are to be defined in advance; and subjective judgment is to be minimized.

In 1973, the Deputy Director (Test and Evaluation) was created within the DDR&E to assume across-the-board responsibility for T&E matters. This has continued until the present except for the period April 1978 through December 1979 when staff responsibilities within OSD were split: the DDT&E had responsibility for DT&E and for overall coordination of T&E, but the ASD, Program Analysis and Evaluation (PA&E) had direct responsibility for OT&E. This policy was reversed in January 1980. Today, the DDT&E responsibilities include:

- Formulating DOD T&E policy and resolving inter-Service problems
- Conducting independent T&E assessments of Service programs to be provided directly to the Secretary of Defense as well as to the DSARC
- Sponsoring joint tests to evaluate operational effectivess of systems which cut across Service lines
- Managing the Major Range and Test Facilities Base (MRTFB).

(3) Test Planning

In the original version of DODD 5000.3, each DOD component was to prepare an overall integrated T&E plan as early as possible but there was no requirement for the Services to obtain OSD review or approval. In 1975, the plan was identified as a separate formal document, called a TEMP, but the requirements for review and approval were still ambiguous, partly because of Service objections to what they believed was additional OSD control of their programs. The ambiguity continued through 1978 when the Services were required to submit their TEMPs to DDT&E only for coordination/review, while at the same time approval somehow was required prior to milestone II. Full clarification occurred in the latest revision of 5000.3: TEMPs must now be submitted to OSD for approal prior to milestone I, and detailed guidelines for TEMP preparation are included in the directive. Changes to T&E plans are to be reported promptly to DDT&E.

(4) Development T&E

DT&E has changed little over the last decade. The objective of DT&E continues to be to demonstrate that the design process is complete, that risks have been minimized, and that the system will meet its design specifications. DT&E is normally planned, conducted, and monitored by the developing agency and has specific objectives for each acquisition phase. Earlier DT&E has been stressed in accordance with the introduction of Phase 0 as has post-milestone III testing with regard to evaluating proposed system modifications. The importance of evaluating software and reliability, availablility, and maintainability (RAM) during DT has recently been emphasized, and the developing agencies have been asked to support OT&E objectives to the maximum extent.

(5) Operational T&E

The original DODD 5000.3 stated that it was important to complete initial operational test and evaluation (IOT&E) before the major production decision; that IOT&E was to be conducted in as realistic an environment as possible; and that the system was to be operated and maintained by typical Service personnel. Subsequent revisions required earlier OT&E and were more specific with regard to the objectives of OT&E during each acquisition phase. The type of hardware to be employed for IOT&E has gradually changed from pilot production, to preproduction prototype, to full-scale engineering models as long as they are reasonably representative of the production configuration. This trend is somewhat inconsistent with the acquisition policy that developed during the same period which allowed limited production for OT&E after milestone II.

(6) Independent Evaluation

Independent evaluation was one of the key recommendations of the Blue Ribbon Defense Panel and was perhaps the major innovation of the T&E policies introduced with 5000.3. Originally, each service was required to establish one (or a limited number of) field agencies responsible for OT&E. These agencies were separate and distinct from the developing/procuring command and also were able to establish strong headquarters staff focal points to assist in conducting required test and evaluation. In 1975, the policy was changed so that only one field agency was permitted. In addition, it had to be separate from the using command as well as the developing/procuring command. The theory was that both users and developers had their own unique biases which would tend to affect their evaluation.

(7) Combined DT and OT

The practice of combining DT and OT was officially proscribed during the early 1970s. The management philosophy then in effect was that effective OT required exercising the system over the full range of expected operating conditions, in conjunction with other systems with which it must function, and under the most realistic conditions possible. It was believed that limited and carefully controlled environmental conditions, operation and maintenance by technician-level personnel, and added instrumentation

commonly imposed by the development tester prevented a true operational test and contaminated the test results. While separate DT and OT are still thought to be most desirable, time and cost pressures have gradually altered the official policy. In 1978, DT and OT were generally to be conducted separately but could be combined when clearly identified cost/schedule benefits or risk is acceptable. The normal practice, however, was for the OT to be conducted independently in support of a production decision. The 1979 revision to 5000.3 states that DT and OT may be combined when appropriate but that, in such cases, the OT&E agency will be a full partner and that OT credibility will be maintained. Final testing prior to milestone III will now emphasize operational testing.

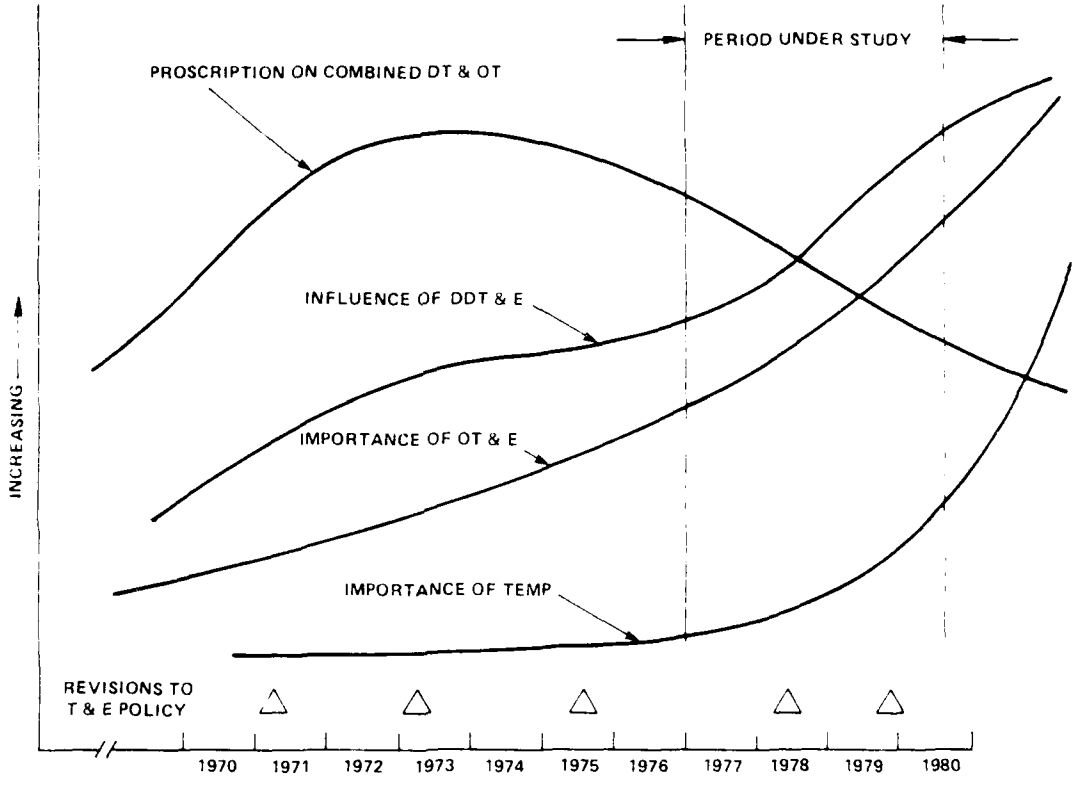
(8) Summary

Exhibit III-3 shows selected trends in T&E policy in a fashion similar to that in which Exhibit III-2 showed trends in overall acquisition policy. Trends in T&E policy may be summarized as follows:

- The most significant trend is the growing influence of the DDT&E in DSARC decision-making. Part of this influence derives from clearer authority granted through changes in high-level acquisition directives and instruction such as 5000.1, 5000.2, and 5000.3. There also is an indication that the DSARC principals may have come to respect DDT&E independent assessment of test results and test planning and have, in fact, become dependent for DDT&E identification of performance and operational suitability issues.

  Similarly, the Services in recent years have recognized the importance of early and continuous communication with the DDT&E, not only because of possible effects on their programs, but also because of their belief that a better test program and a better product will result. The increased slope of the DDT&E influence curve after about 1978 reflects the vigorous activites of the office to institute stricter control and use of the TEMP, to emphasize the importance of assessing operational suitability early in the life cycle, and to increase the use on analytical techniques. The office is presently regarded as both competent and fair by OSD and the Services alike.

EXHIBIT III-?
Trends in T&E Policy



- The importance of OT&E has been steadily increasing throughout the decade. This is consistent with the emphasis on operational suitability noted earlier. The present view is that, short of actual warfare, operational testing is the only legitimate means of assessing how a system will perform in its operational environment. The trend acknowledges the difficulty in establishing technical system specifications and the fact that meeting them (generally as evaluated in DT) does not necessarily mean the system will perform its intended function. It should be noted, however, that policy has backed away from requiring OT&E on production or preproduction hardware prior to the milestone III review and that combining DT and OT is now accepted under some circumstances as a means of accelerating the acquisition process (see below).

- Running somewhat counter to the emphasis on OT&E is an apparent softening of the attitude on the deleterious effects of combining DT and OT. However, this trend does parallel the emerging philosophy on concurrency, which, in fact, reflects a pervasive trend throughout acquisition policy: that absolute or cookbook approaches are not optimally effective; rather, that the process should be tailored to suit the individual circumstances. In this case, it an acknowledgement that T&E activities can be reduced when risks permit.

- While the concept of formal T&E planning in the context of decentralized operations was introduced in the 1970-1971 time frame, the TEMP did not become a significant planning tool until the end of the 1970s. The explanation lies in the initial reluctance of the Services to allow OSD such a detailed role in the management of their programs. This also explains the ambiguity of early policy statements on the purpose of the document. According to an OSD study, until early 1979 the TEMP was treated casually even by DDT&E. Subsequent recognition that it needed to be elevated in stature and content led to the changes in 5000.3 discussed earlier. Now, the status of TEMPs are tracked much more carefully, and DDT&E maintains statistics on currency, approval, revisions, etc.

Although not a trend, an important event in the evolution of T&E policy occurred in 1974 when changes were made to the procedures by which the major ranges and test facilities were reimbursed for services provided to individual users. Previously, the ranges had been predominantly institutionally funded from their own Operations and Maintenance (O&M) lines of the DOD budget. Most services were provided at no cost to the specific programs. Now, in accordance with the Uniform Funding Policy, users are required to reimburse the test activity for all direct costs that can be associated with a particular program; only general support costs are funded by O&M accounts.

While the Uniform Funding Policy does not change the overall cost of testing within DOD, it does shift the burden from those preparing and justifying the O&M budget to those advocating specific weapon systems. In general, the increased cost of doing business from the point of view of the program manager has not been reflected in increases to program budgets. This has added to the difficulty in conducting thorough, comprehensive test programs and has encouraged PMs to make trade-offs between the cost of unplanned development problems and the cost of planned testing.

3. HISTORICAL PERSPECTIVE

Overviews of acquisition reform and evolution of T&E policy provide one perspective for evaluating the DSARC decisions that took place during the period of interest. Another perspective can be gained by putting policy events along with other events in a time-phased display called a diachronic chart. Such a chart for the period 1970 through 1981 is shown in Appendix D. It shows the following:

- The names and terms of senior Administration officials
- Landmark policy events, such as the President's Blue Ribbon Defense Panel and the issuance of OMB Circular A-109
- Acquisition- and T&E-related directives and instructions
- Significant elements of the OSD decisionmaking process particularly as it relates to T&E
- Selected T&E studies conducted by Government agencies and outside task forces

- GAO reports related to T&E in general and the effectiveness of DOD system acquisition decision practices
- Events relating to Congressional involvement in T&E aspects of the acquisition process
- Selected defense-related world events
- The names and terms of the DSARC principals.

The diachronoic chart is best used as a reference on which to overlay other data. However, some observations can be made about the information contained in the chart itself beyond those already discussed on acquisition and T&E policy.

The U.S. political process causes a turnover in key personnel at 4- or 8-year intervals, notwithstanding other changes. This means that the management philosophy or emphasis espoused by one administration may not continue into, or may even be reversed by, the next. It is obvious that establishing a stable framework for acquisition decisionmaking under these circumstances is difficult. Furthermore, the turnover in DSARC principals from one milestone to the next on a given program tends to make each review unique rather than an update of previous decisions and strategies.* The importance of the DOD career civil servants in providing continuity under these circumstances should not be underestimated.

Another significant point reflected in the chart is that Congressional and GAO interest in detailed acquisition issues and decisions, especially as they relate to T&E, has risen steadily over the decade. In 1971, Congress passed Public Law 92-156 which required the Secretary of Defense to submit to Congress each year, beginning in FY73, "... data on operational testing and evaluation for each weapon system for which funds for procurement are requested (other then funds requested only for the procurement of units for operational testing and evaluation

---

* Data from the final report on Weapons System Costs from the DSARC Cost Reduction Working Group (the "Little Four"), December 19, 1972, indicate that the participation by the same DSARC principal on 11 repeat or successive DSARCs held between September 1969 and June 1971 ranged from 64 to less than 10 percent. This inconsistency in attendance was caused by substitution for the DSARC principal rather than changes in appointment.

and/or long lead-time items)." These data, and other information, are included in Congressional data sheets which are submitted each year with the President's budget.

In 1975, Congress enacted Public Law 94-106 which established the Selected Acquisition Report (SAR) as a legal reporting requirement of the DOD for designated systems. Originally, the SARs were instituted as an internal DOD mechanism for keeping the ASD Comptroller apprised of the progress of selected acquisitions and comparing this progress with technical, schedule, and cost plans. In 1969, in response to a request from the Chairman of the Senate Armed Services Committee, the Secretary of Defense agreed to provide this information to Congress. Today, the SARs are used primarily by both the Senate and House Armed Services and Appropriation Committees, the House and Senate Budget Committees, and the House Government Operations Committee. The SAR is also sent to the GAO, OMB, and other agencies.

According to unpublished data assembled by the Director, Management Information Control and Analysis, ASD(C), Congress indeed is playing a stronger role in the acquisition process as evidenced by almost every conceivable measure, such as total number of employees, professional staff members on defense-related committees, witnesses testifying before committees, line items in the budget, and number of pages in the defense authorization and appropriation committee reports. In addition, the GAO since fiscal year 1960 has nearly tripled the number of defense-related reports they send to Congress each year. As seen in the diachronic chart, GAO reports on T&E, acquisition decisionmaking, and relations with Congress have increased steadily since 1972. Not shown on the chart are the weapon-specific reports prepared by the GAO at the beginning of each calendar year, although these reports are shown on "mini-diachronics" for each program discussed in Chapter V.

Another important aspect of the diachronic chart is its ability to show the connectivity between various events. Sometimes this is obvious, as when a revision of a directive follows its original publication. Other cases are more subtle, such as the effect of a major study or report of a commission on subsequent revisions to policy. This process also shows that sometimes the policy change actually occurred before the landmark event which presumably caused it; that is, the landmark event was actually confirming what had generally become accepted rather

than breaking new ground in itself. Where connectivity between events can be inferred and is practical to display, it is shown as a dotted line.

* * * * *

The decision environment provides a backdrop against which to analyze the 38 DSARC II and III reviews held during the period under study. The results of these DSARCs are described in a macroscopic sense in the next chapter.

IV. DSARC DECISIONS

IV. DSARC DECISIONS

There have been nearly 250 DSARCs since the process was instituted in 1969. As shown in Exhibit IV-1, there are about 15 to 20 DSARCs held each year, with a peak of 30 in 1975. This chapter addresses the 38 DSARC II and III reviews held between January 1977 and August 1980. It includes:

- An overview of the milestone II and III reviews
- The identification and characterization of the decisions made by the DSARC
- An analysis of trends and relationships exhibited
- Identification of the apparent reasons for each decision
- A rationale for selection of programs for further study.

1. OVERVIEW

As shown in Exhibit IV-2, there were 38 milestone II and III reviews which addressed 34 different weapon systems during the period under study. The dotted lines indicate that the air-launched cruise missile (ALCM) program had both milestone II and III reviews, the F-16 had both IIIA and IIIB reviews, and the high-speed antiradiation missile (HARM) and MX programs had IIA and IIB reviews. Follow-on reviews were planned for the major caliber lightweight gun (MCLWG), Roland, and XM-1 programs, but they did not occur during this period.

The table below shows the distribution of reviews by Service and milestone.

SERVICE	MILESTONE II REVIEWS	MILESTONE III REVIEWS	TOTAL
ARMY	3	9	12
NAVY	5	5	10
AIR FORCE	6	8	14
JOINT	2	–	2
TOTAL	16	22	38

EXHIBIT IV-1
DSARC History

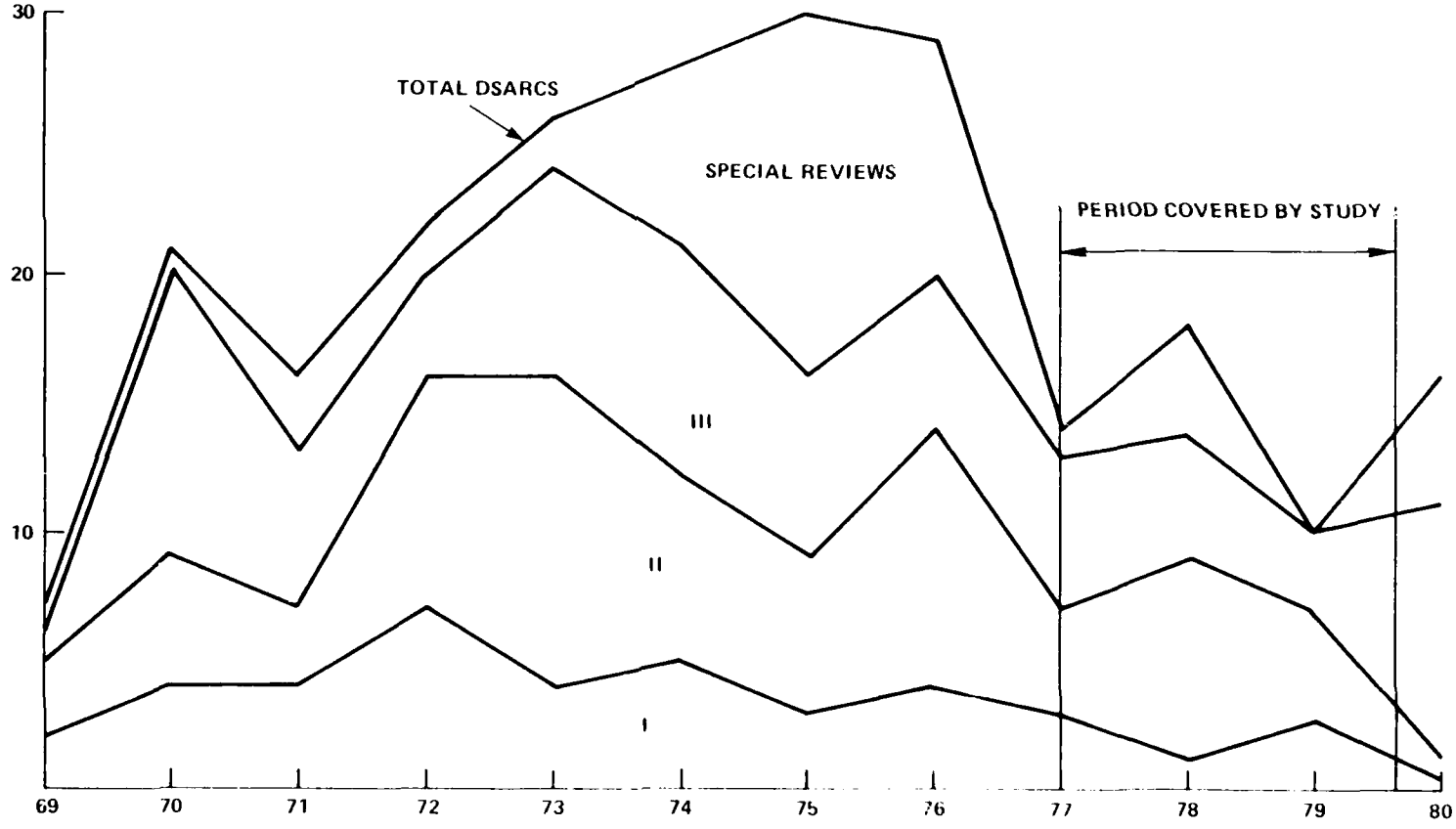
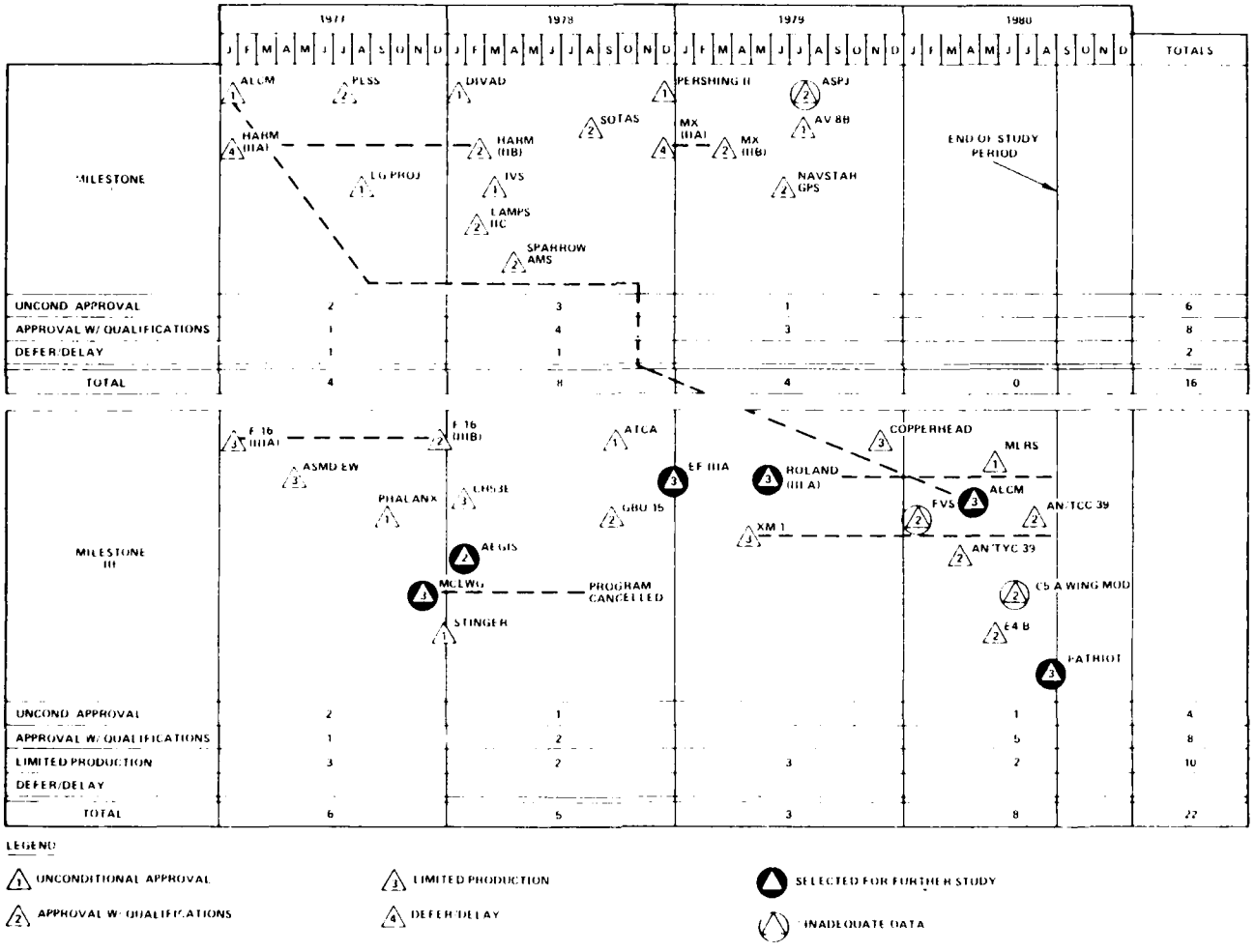


EXHIBIT IV-2
DSARC Decisions
January 1977 - August 1980



LEGEND

1 UNCONDITIONAL APPROVAL
2 APPROVAL W/ QUALIFICATIONS
3 LIMITED PRODUCTION
4 DEFER/DELAY
SELECTED FOR FURTHER STUDY
INADEQUATE DATA

Of the 38 DSARCs, 12 addressed Army programs, 10 Navy programs, 14 Air Force programs, and 2 joint programs. The joint programs included the airborne self projection-jammer (Navy/AF) and the NAVSTAR global positioning system (GPS) program (Navy/AF/Army). The table also shows that 16 of the 38 DSARCs were milestone II decisions and the remaining 22 were milestone III decisions.

2. CHARACTERIZATION

In general, DSARC decisions fall into three basic categories: unconditional approval to proceed to the next phase, approval to proceed to the next phase with certain qualifications, and disapproval (cancellation) or delay in proceeding to the next phase pending resolution of technical or other issues raised by the DSARC. In the case of the production decision (milestone III), a fourth category, limited production, is often employed and is helpful in understanding the concerns of the DSARC principals and in analyzing trends in acquisition policy.

(1) Milestone II Reviews

The 16 milestone II reviews are shown in the table below by Service and type of decision.

DECISIONS / SERVICE	UNCONDITIONAL APPROVAL	APPROVAL W/QUALIFICATIONS	DEFER/ DELAY	TOTAL
ARMY	2	1	0	3
NAVY	2	2	1	5
AIR FORCE	2	3	1	6
JOINT	0	2	0	2
TOTAL	6	8	2	16

Six DSARCs (38 percent) gave unconditional approval to proceed to FSD: ALCM, AV-8B, laser guided projectile, DIVAD, IUS, and Pershing II. These systems had completed their test programs for the demonstration and validation phase successfully and had met all thresholds substantially. No significant T&E issues had been identified and no additional aadvanced development (AD) testing was recommended. Eight (50 percent) of the programs reviewed at milestone II received approval to proceed to FSD with certain qualifications.

In these cases, the test data and other information provided to the DSARC were regarded as adequate to support entry into FSD if additional steps were taken to reduce risks; i.e., the risks involved in proceeding were considered to be acceptable in the context of the entire body of existing data or when weighed against an alternative to delay or cancel the program. The qualifications imposed were usually either to correct all identified decifiencies and demonstrate the successful correction by additional testing at a later stage or to increase confidence in certain specific areas by providing additional relevant test data. In two cases (12 percent), HARM and MX, it was decided to delay the decision to proceed to FSD; i.e., a decision was made to remain in the advanced development phase until either specific problems were resolved as demonstrated by additional testing or the program was redirected, depending on the case under consideration.

It is interesting to observe that while the Services' programs were fairly evenly distributed over the decision categories, neither of the two joint programs received unconditional approval.

(2) Milestone III Reviews

The 22 Milestone III decisions were distributed by Service and type of decision as shown below.

DECISION / SERVICE	UNCONDITIONAL APPROVAL	FULL-SCALE PRODUCTION W/Qualifications	LIMITED PRODUCTION W/Qualifications	DEFER/ DELAY	TOTAL
ARMY	2	3	4	–	9
NAVY	1	1	3	–	5
AIR FORCE	1	4	3	–	8
JOINT	–	–	–	–	–
TOTAL	4	8	10	0	22

Four systems (18 percent) had completed their test programs for the FSD phase successfully and met all thresholds: Phalanx, Stinger, ATCA, and MLRS. No significant T&E issues were identified and test data appeared to support an unqualified decision to proceed to rate production ("maturation phase" in the case of MLRS). Eight systems (36 percent) received approval

to proceed to rate production with certain qualifications; i.e., proceeding to rate production was made conditional or dependent on resolving specific problems, usually in parallel with the production effort and generally not requiring another formal DSARC before proceeding. Nearly half (45 percent) of all programs at milestone III received approval for only limited production. For these programs, the risks of committing to production were regarded to be too high without an extended evaluation of limited production hardware. A favorable full-scale production decision was postponed until additional testing had demonstrated either the successful correction of identified deficiencies or the achievement of thresholds which had not yet been tested.

There were no milestone III reviews that resulted in outright cancellation of ongoing programs or even in denial of proceeding to production in some manner.

(3) Summary

Overall, nearly three-quarters (74 percent) of the milestone II and III reviews resulted in less than unconditional approval: 10 out of 16 (62 percent) milestone IIs, and 18 out of 22 (82 percent) milestone IIIs. The totals for each decision category are shown by year at the bottom of each section of Exhibit IV-2.

3. ANALYSIS OF TRENDS AND RELATIONSHIPS

A plot of the DSARC decisions is shown in Exhibit IV-3. This graph shows the total number of DSARC II and III reviews held each year and the number that resulted in unconditional approval. If the data are normalized, as shown in Exhibit IV-4, it can be seen that the percentage of unconditional approvals is decreasing almost linearly from about 40 percent in 1977 to less than 15 percent in the first 8 months of 1980.

These data could be interpreted as reflecting the increased emphasis on operational effectiveness and suitability that occurred during this time frame and, perhaps, a more cautious attitude by defense decisionmakers regarding deploying systems that had not met their operational suitability goals. However, it is also possible that this trend results from the "bow wave" phenomena--programming for more systems than can be afforded. A 1977 Defense Science Board (DSB) task force report on the defense acquisition cycle estimated that the bow wave averaged 30%

EXHIBIT IV-3
DSARC Decisions vs. Time

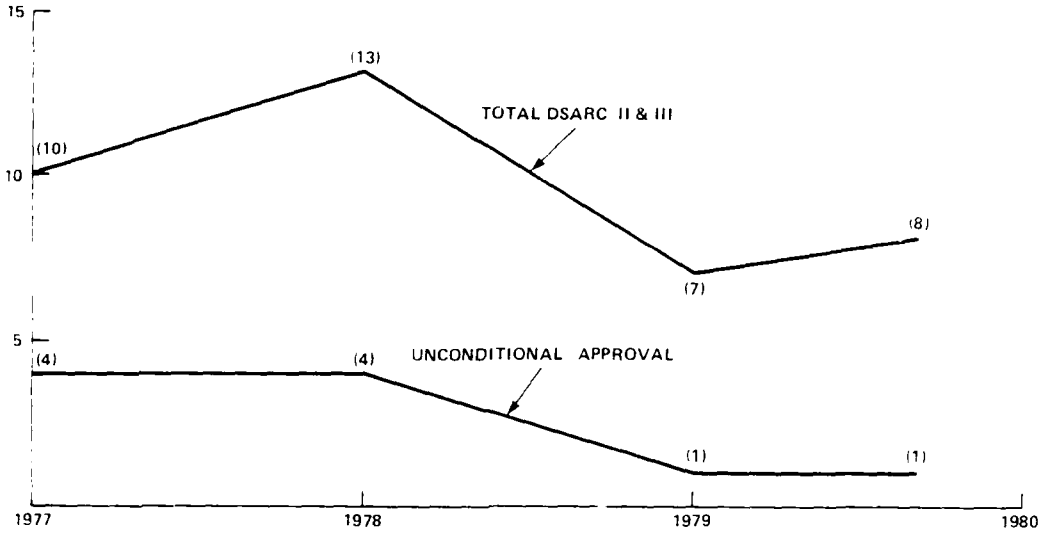
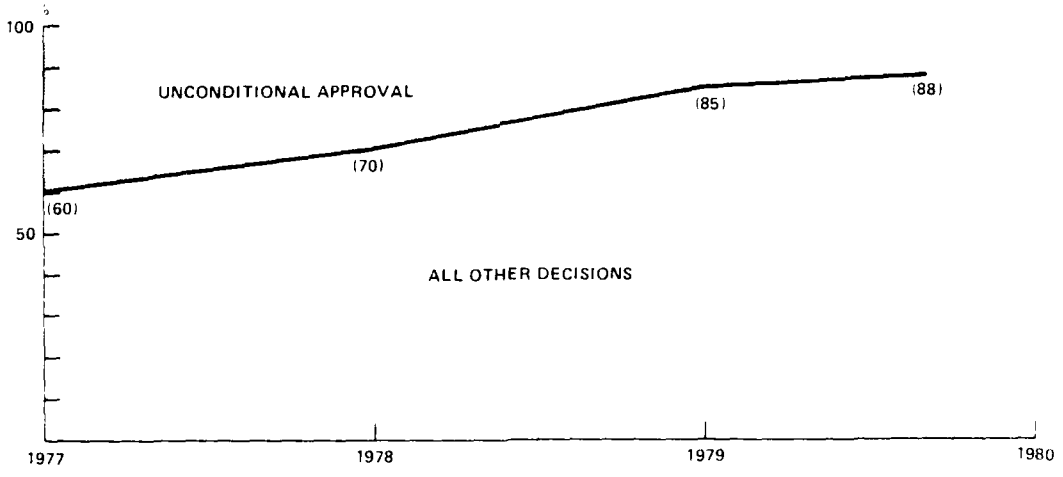


EXHIBIT IV-4
DSARC Decisions vs. Time
(Normalized)



rather than the budget in the procurement accounts. The 1970 DSB task force report on reducing the unit cost of equipment concluded that maintaining the inventory with the current procurement budget would require roughly a 40% reduction in average unit cost. The decisionmakers' perception of the bow wave was, and continues to be, a strong influence on any decision to proceed aggressively into the next phase of acquisition.

Another important question is whether the DSARC viewed milestones II and III differently with respect to approving a system for the next phase. Most members of the acquisition community believe that the most critical decisions are made at milestone 0 'when a formal system acquisition is actually begun and DOD commits itself to satisfying a need' and at milestone II 'when the DOD commits itself to full-scale development and states an intention to produce and deploy the system'. The importance of milestone II is due to the sense of commitment to the program and the lack of opportunity for change which exists later at the time of the production decision. By milestone III, the program has acquired a broad constituency with very specific vested interests in its continuation. Furthermore, by milestone III, the time required to pursue an alternative course is almost always unacceptable in light of the threat. The DSB in 1977 concluded that if test issues were raised at milestone III, the DOD was faced with "... equally unpalatable alternatives:

> ... either cancel the program which wastes the investment already made and denies the product to the field; order a major rework which is enormously expensive after FSD has been completed; or bite the bullet and go ahead with production anyway. The pressure to do the last is enormous ..."

In addition, Mr. Dale Church, Deputy Under Secretary of Defense for Acquisition Policy, testified before the Congress (in November 1979) that:

> The key decision, at Milestone II to begin Full-Scale Engineering, is very significant because it represents a decision to greatly increase the resources expended. Thus, a decision to proceed at this point must be tantamount to commitment to produce the system if it achieves its expected performance.

This implies that OSD and the DSARC should be more critical in their evaluation at milestone II and that, theoretically at least, unconditional approval should be less frequent at that time.

In fact, however, [illegible] there was a greater percentage of unconditional approvals at milestone II (6 percent) than at milestone III (4 percent). On the basis of a chi-square test of independence (see Appendix D), a hypothesis that the percentage of unconditional decisions is statistically the same at each milestone cannot be rejected; however, it seems clear that the preponderance of analysis and the intensity of high-level management interest do not exist at milestone II as they should.

One explanation for this phenomenon is that the highest level DSARC decisionmakers are political appointees with rather short terms of office compared to the length of the budget cycle or the lead time in acquiring new systems. This results in their primary interest being focused on the near-term cost impact of decisions such as those made at milestone III, rather than the longer-range implications of decisions made at milestone II. This built-in political shortsightedness is recognized by many, but there is no clear solution as long as the R&D production cycle is greater than the political cycle.

4. APPARENT REASONS FOR DSARC DECISIONS

Having characterized the DSARCs by decision type (unconditional approval, approval with qualifications, etc.), the investigation can now attempt to identify the apparent reasons for each decision. The word apparent is used to reflect the fact that it is not always obvious what the motivations of the DSARC principals were on the basis of the formal documentation.

Using the syntax described in Chapter II, reasons for less than unconditional approval can be grouped in need, affordability, maturity, management plans, as well as other external influences. Maturity is further subdivided into failure to meet thresholds and inadequate test data. Finally, these categories are subdivided into more specific reasons, as shown below:

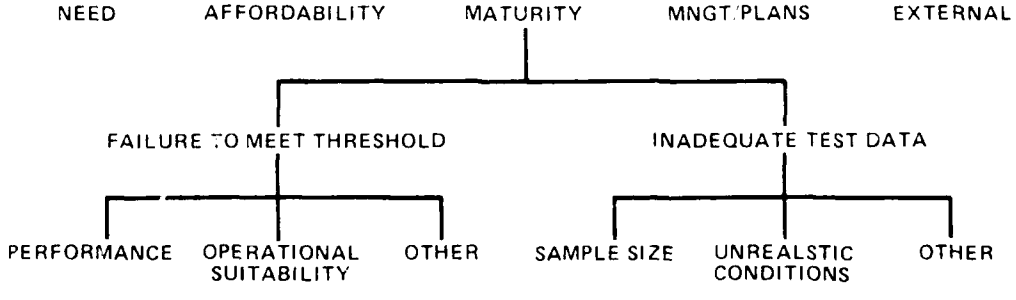


Exhibit IV-5 lists in alphanetical order each weapon system that had a milestone II or III review during the period under study. In addition, the DSARC type and date, the decision type, a narrative description of T&E issues contained in the DSARC documentation, and an indication of the apparent reason(s) for less than unconditional approval (where applicable) are shown.

Exhibit IV-6 shows the frequency of apparent reasons for conditional decisions as derived from Exhibit IV-5. Inspection of these data shows that failure to meet thresholds is the predominant, apparent reason for conditional decisions. Twenty-one of the 28 DSARCs which restricted the pace of the program or ordered additional T&E did so because the system failed to meet one or more of its performance or operational suitability thresholds. This included 5 out of 10 DSARC IIs (50 percent) and 16 out of 18 DSARC IIIs (89 percent). This would imply that the performance specifications were set unrealistically high, that the reviews were being held too early in the acquisition cycle to allow for the desired degree of system maturity, or that the DSARC failed to appreciate the developmental nature of the systems under review.

Inadequate T&E data appeared to be a reason for less than unconditional approval in only nine cases: one DSARC II and eight DSARC IIIs. These included three Army programs (the Roland and Patriot missile systems and the fighting vehicle system), two Navy programs (the MCLWG and Aegis weapon system), and three Air Force systems (the ALCM, the EF-111A tactical jamming system, and the C-5A wing modification).

As shown in the note to Exhibit IV-6, the apparent reasons occurred singly and in combination. The most frequent combination was a breach of performance threshold and other reasons; next was threshold breach plus inadequate T&E data; followed by threshold breach alone and other reasons alone. Inadequate T&E data was the sole reason for conditional decisions in only two cases, the C-5A and the fighting vehicle system (FVS). Interestingly, both of these (and the Aegis review) resulted in a full production release with qualifications rather than only limited production authority granted the other programs.

As shown in Exhibit IV-7, the percentage of conditional decisions that can be attributed, at least in part, to inadequate T&E data has risen steadily since 1977. Inadequate T&E data appeared to be a reason for imposing some form of constraint on more than half of the programs which had such decisions in the first 8 months of 1980. However, it is inappropriate to extrapolate this curve into the future since it reflects policies that were in effect 5 or more years ago.

EXHIBIT IV-5

Summary of Milestone Decisions

SYSTEM	MILESTONE	DATE	DECISION	T&E ISSUES	NEED	AFFORD ABILITY	MATURITY: FAILURE TO MEET THRESHOLD — Performance	MATURITY: FAILURE TO MEET THRESHOLD — Op Suit incl R&M	MATURITY: FAILURE TO MEET THRESHOLD — Other	MATURITY: INADEQUATE TEST DATA — Sample Size	MATURITY: INADEQUATE TEST DATA — Unrealistic Conditions	MATURITY: INADEQUATE TEST DATA — Other	MANAGEMENT & PLANS	OTHER EXTERNAL
AEGIS	III	19 Jan 78	FSP w Qualifications	Testing generally going well but only 25% of combined DT OT completed as of 1/78 inadequate to assess system effectiveness and suitability Software a critical T&E issue software development only 45% complete as of 1/78 Additional testing of fully integrated system required to increase system understanding			●				●	Absence of key subsystems		
ALCM	II	6 Jan 77	FSD Unconditional	SECDEF directed that JCMPO be established and that Joint T&E program be conducted for FSED										
	III	17 April 80	LP w Qualifications	In flight test program several deficiencies noted more extensive DT and OT required because A Problems identified in FSED especially in reliability B Some subsystems still undergoing development testing not completed in FSED			●	●		●	●	Instrumentation problems absence of key subsystems		
AMTCC 39	III	8 Jul 80	FSP w Qualifications	Appropriate test plan for OSD approval should be submitted prior to contract award JCS representative disagrees with this condition									●	
AMTYC 39	III	25 Mar 80	FSP w Qualifications	Additional testing to be performed by OTEA to verify correction of deficiencies in reliability and maintainability				●						
ASMD EW	III	12 April 77	LP w/Qualifications	Testing revealed several performance deficiencies must judged to be software related System improvements additional testing required before full production decision			●							
ASPJ	II	17 Jul 79	FSD w Qualifications	Navy directed to increase the test data base to achieve a high level of confidence in reliability maintainability				●		●				
ATCA	III	Sept 78	FSP Unconditional	Program well planned technical and cost risks low No T&E issues										
AV 8B	II	18 Jul 79	FSD Unconditional	It was recommended to conduct some form of shipboard suitability trials Concern was expressed about limited flight test experience and its impact on meeting DSARC III R&M thresholds										
CSA WING MOD	III	June 80	FSP w Qualifications	Actual fabrication deferred until additional fatigue testing is performed TEMP revision required								Complete lifetime fatigue test	●	
CH 53E	III	31 Jan 78	LP w Qualifications	Performance and reliability maintainability thresholds not met in test program Improvement program additional testing required			●	●						

KEY FSD Full Scale Development
FSP Full Scale Production
LP Limited Production

NOTES
1 DSARC authority delegated to AT/SARC held 10 Aug 1978
2 No formal DSARC held

EXHIBIT IV-[illegible] (Continued)

SYSTEM	MILESTONE	DATE	DECISION	T&E ISSUES	NEED	AFFORDABILITY	MATURITY: FAILURE TO MEET THRESHOLD: Performance	MATURITY: FAILURE TO MEET THRESHOLD: Op Suit incl RAM	MATURITY: FAILURE TO MEET THRESHOLD: Other	MATURITY: INADEQUATE TEST DATA: Safety Surv	MATURITY: INADEQUATE TEST DATA: Operating Conditions	MATURITY: INADEQUATE TEST DATA: Other	MANAGEMENT & PLANS	OTHER EXTERNAL
COPPERHEAD	III	6 Nov [illegible]	LP w Qualifications	Army directed to establish a reliability growth program, perform additional testing to demonstrate reliability improvement. Also [illegible] CT live fire training program required		●		●						
DIVAD	[illegible]	[illegible] Jan 78	FSD Unconditional	Recommend an austere prototype test program to demonstrate system integration										
F-15	[illegible]	[illegible] Mar [illegible]	[illegible] [illegible]	Testing for EMP hardness verification is a major issue; testing to be performed on TRESTLE facility. Software, comm processing [illegible] also [illegible] to be tested		●	●	●					●	
[illegible]	[illegible]	[illegible] Oct [illegible]	LP w Qualifications	Significant problems with operational suitability, subsystem effectiveness and lack of test facilities to test realistically [illegible] environments. T&E assessment of [illegible] did not [illegible] decision				●			●			
FIGHTING VEHICLE SYSTEM	[illegible]	[illegible]	[illegible] FSP w Qualifications	[illegible] to expedited to [illegible] [illegible] Reliability [illegible] evaluated. [illegible] testing required								[illegible]		
[illegible]	[illegible]	[illegible]	[illegible] w Qualifications	[illegible] expressed regarding optimistic test schedule				●					●	
	[illegible]	[illegible]	FSD w Qualifications	Additional operational testing required to [illegible] the [illegible] achieve the desired operational suitability. Test results [illegible] RAM				●						
[illegible]		[illegible]	[illegible] w Qualifications	[illegible] performance threshold [illegible] additional testing [illegible] required [illegible]				●					●	
HARM	[illegible]	[illegible]	[illegible]	System failed to meet technical thresholds at high frequencies. Maintainability was deficient. Concern was expressed about [illegible]		●	●							
	[illegible]	[illegible] Feb [illegible]	FSD w Qualifications	Test program to be modified; increase amount of [illegible] to include assessment of several performance parameters			●						●	
		[illegible] Mar [illegible]	[illegible]	Development testing on prototype components looks good, but [illegible] reliability will not be demonstrated before [illegible]										
[illegible]	[illegible]	[illegible]	FSD w Qualifications	Affordability problems resulted in scaleback of [illegible] testing. Navy [illegible] additional attention to availability, reliability, maintainability, and logistics support				●						
[illegible]	[illegible]	[illegible] Nov	LP w Qualifications	[illegible] system testing [illegible] conducted by the [illegible] of MIL [illegible] Additional testing required for system integration [illegible] of operational capabilities and RAM improvement				●				[illegible]		

[illegible]

EXHIBIT IV-5 (Continued)

SYSTEM	MILESTONE	DATE	DECISION	T&E ISSUES	NEED	AFFORDABILITY	MATURITY: FAILURE TO MEET THRESHOLD: Performance	MATURITY: FAILURE TO MEET THRESHOLD: incl R&M Op Suit	MATURITY: FAILURE TO MEET THRESHOLD: Other	MATURITY: INADEQUATE TEST DATA: Sample Size	MATURITY: INADEQUATE TEST DATA: Unrealistic Conditions	MATURITY: INADEQUATE TEST DATA: Other	MANAGEMENT & PLANS	OTHER EXTERNAL
MLRS	III	27 May 80	Maturation Phase approved (equivalent to FSP Unconditional)	Army directed to put special effort into testing of built in test of other maintenance features										
MX	IIA	5 Dec 78	Program redirection	Program was required to include Air mobility study									●	
	IIB	31 Mar 79	FSD w/Qualifications	Test plan developed by Air Force but T&E assessment not yet available from ODT&E									●	Higher Auth. Required Presidential Decision
NAVSTAR GPS	II	5 Jun 79	FSD w/Qualifications	Affordability issue resulted in recommendation to build only a Qualifying Test Vehicle thereby eliminating the phase II prototype satellite		●								
PATRIOT	III	21 Aug 80	LP w/Qualifications	Testing to date has demonstrated deficiencies in performance in ECM; poor reliability maintainability unsatisfactory mobility and immature software. Additional testing to be performed to demonstrate achievement of performance and RAM thresholds before full production approval			●	●		●	●	Maintainability Could Not Be Evaluated		
PERSHING II	II	21 Dec 78	FSD Unconditionally	Advanced development testing at WSMR successfully demonstrated feasibility of subsystems. T&E assessment supports transition to FSED										
PHALANX	III	20 Sep 77	FSP Unconditionally	Extensive testing performed thresholds achieved no significant T&E problems/issues identified										
PLSS	II	26 Jul 77	FSD w/Qualifications	Several areas of TEMP require improvement									●	
PROJECTILE II G	II	23 Aug 77	FSD Unconditionally	AD testing demonstrated operational feasibility however design changes in process will require testing during FSED (ASD (PA&E) judged risks to be higher than usual)										
ROLAND	III	31 May 79	LP w/Qualifications	Test results showed significant problems in reliability maintainability testing in other areas incomplete as of 5/79. A RAM improvement program recommended to be demonstrated in testing before a full production decision. Mod. of test plan			●	●			●	Maintainability Not Tested	●	
SOTAS	II	4 Aug 78	FSD w/Qualifications	Army directed to program DT and OT test time during FSED for compliance with logistics, reliability maintainability and availability goals with high confidence				●					●	
SPARROW AIMS	II	27 Apr 78	FSD w/Qualifications	Decision to authorize initial production of long lead missile components contingent upon achievement of thresholds to be demonstrated by testing and successful Joint Technical Evaluation			●							
STINGER	III	Dec 77	FSP Unconditional	Test program very successful no significant T&E issues										
XM1	III	19 Apr 79	LP w/Qualifications	FSED testing demonstrated serious deficiencies in reliability and durability test results did not support a recommendation for full production. Army directed to make RAM D improvements and demonstrate in OT before additional production approval. Revised TEMP required				●					●	

KEY: FSD Full Scale Development
FSP Full Scale Production
LP Limited Production

EXHIBIT IV-4
Reasons for Conditional Decisions

NOTE: REASONS OCCURRED IN COMBINATION AS SHOWN

	M/S II	M/S III	TOTAL
T + D + 0	–	–	0
T + D	1	6	7
T + 0	3	5	8
D + 0	–	–	0
T ONLY	2	4	6
D ONLY	–	2	2
O ONLY	4	1	5
TOTAL	10	18	28

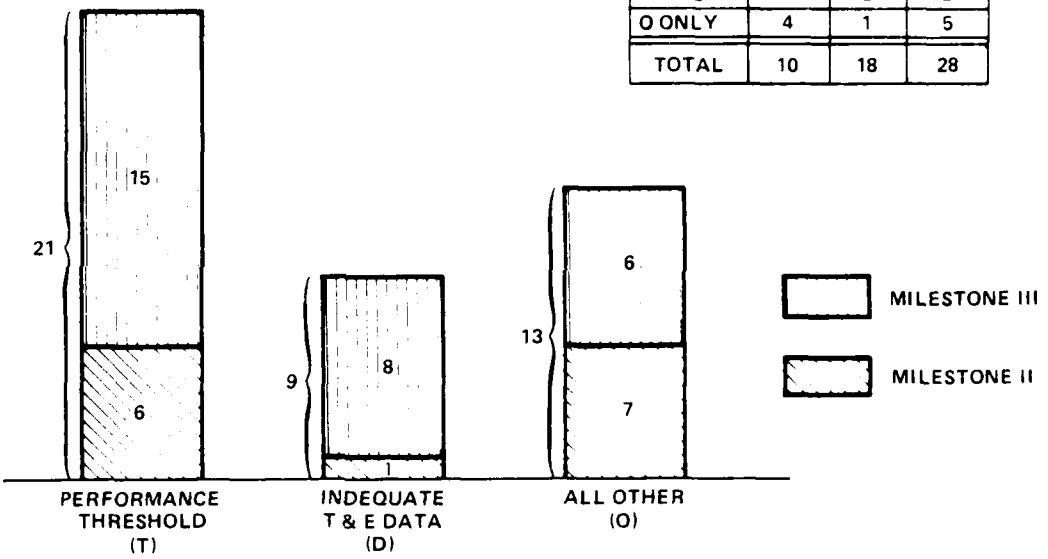
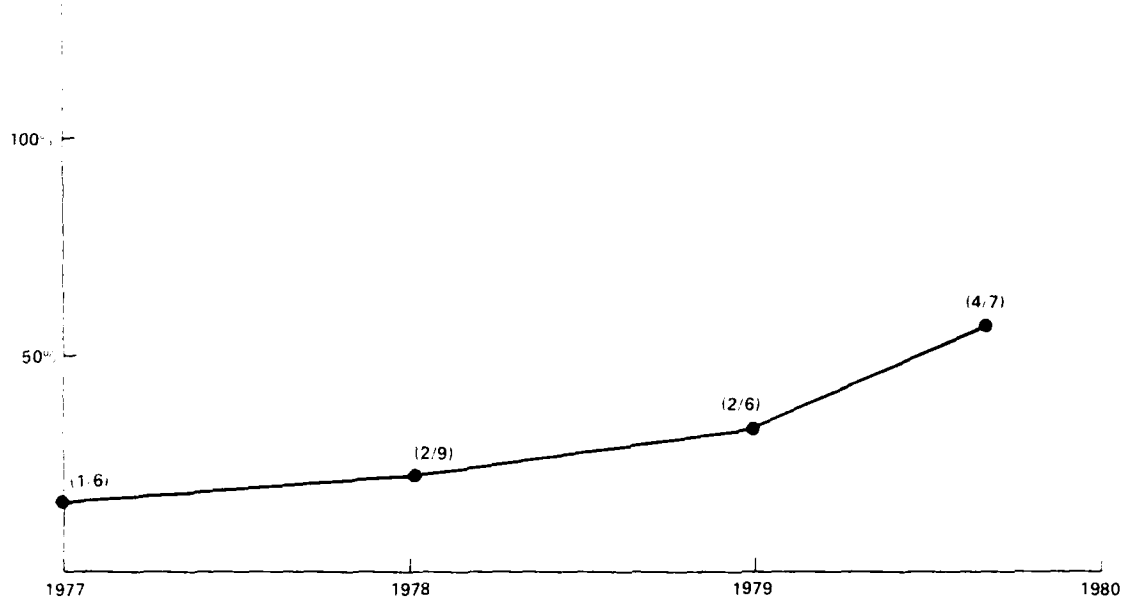


EXHIBIT IV-7
Conditional Decisions and
Inadequate T&E Data



Key: (A/B)

A = Number of conditional decisions attributed, at least in part, to inadequate T&E data.

B = Total number of conditional decisions.

5. SELECTION OF PROGRAMS FOR FURTHER STUDY

On the basis of this initial screening, it was decided to concentrate on those milestone III decisions that appeared to involve inadequate T&E data. Two programs were selected from each Service for further study, as follows:

SYSTEM	DSARC III DATE	DECISION
Army		
Patriot	21 August 1980	Limited production
Roland	31 May 1979	Limited production
Navy		
Aegis	19 January 1978	Full-scale production with qualifications
MCLWG	8 November 1977	Limited production
Air Force		
ALCM	17 April 1980	Limited production
EF-111A	12 December 1978	Limited Production

* * * * *

Chapter V contains a decision profile of each of the selected systems. In order to identify the underlying causes of inadequate T&E data, issues arising from each milestone review are analyzed in the context of programspecific events and of the overall environment in which the decision was made. For this reason, it may be helpful to unfold Appendix D, the diachronic chart, for reference during the review of the decision profiles contained in the next chapter.

V. DECISION PROFILES

V. DECISION PROFILES

This chapter describes the events surrounding the milestone III review of the six programs selected for further study. In each case, there is evidence that the DSARC principals felt there were inadequate T&E data for making an unconditional decision to proceed to full-scale production. This chapter will provide insight on how the DSARC integrated T&E issues with other factors in arriving at their final recommendations.

Each decision profile in this chapter is divided into the following three parts:

- A factual summary of the formal DSARC documentation
- A subjective interpretation of the DSARC decision based on interview results and analysis of program-specific and exogenous factors
- A historical schedule showing program milestones and key events.

In addition, each decision can be put in historical perspective by referring to the diachronic chart contained in Appendix D. The decision profiles are arranged in chronological order, starting with the earliest milestone III review.

1. MAJOR CALIBER LIGHTWEIGHT GUN (MCLWG)

The MCLWG program was initiated in 1963 and progressed under Navy management with several major redirections in scope and configuration, including changes in both caliber and platform. A decision in 1973 to use the DD-963 class as the platform for the MCLWG intensified development efforts and could be considered the actual start of FSD. Exhibit V-1 is a summary of the official documentation surrounding the milestone III review; Exhibit V-2 shows the key events leading to the DSARC.

Interest in the MCLWG program waxed and waned throughout most of its life. It only came to OSD attention as a result of Congressional and GAO interest some 13 years after beginning development. It was the kind of program that had neither high national priority nor a complete commitment by its proponent Service. Although the DSARC approved the Navy's request for limited production in

EXHIBIT V-1
Decision Documentation
Major Caliber Lightweight Gun (MCLWG)

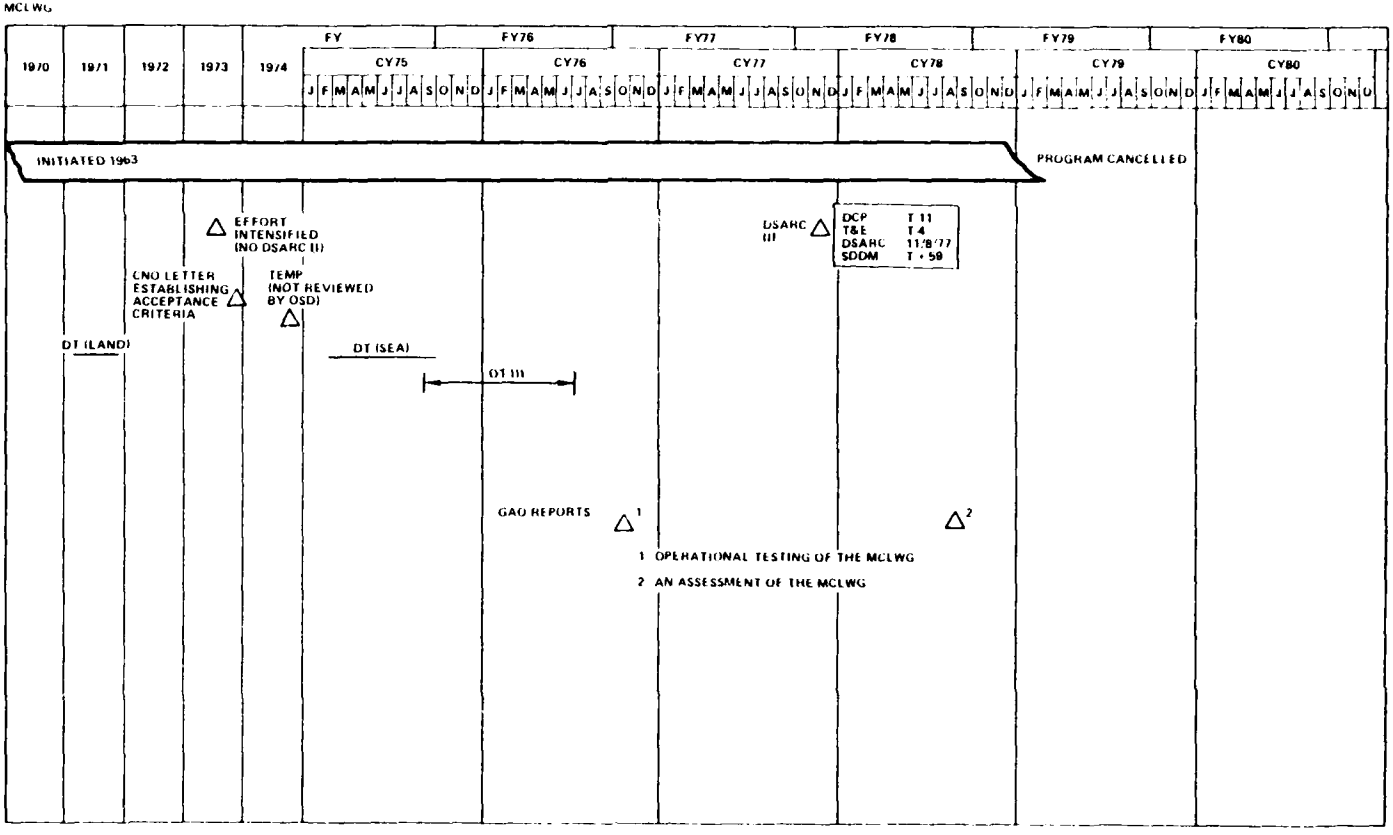
TEMP: TEMP No. 016, dated September 9, 1974 was the basis for testing up to DSARC III. This TEMP was not coordinated with OSD. A new TEMP (No. 118) was in preparation at the time of the DSARC.

PRE-DSARC III DCP: DCP #167, dated 28 October 1977, addressed the issue as to whether production of the MCLWG should be approved. It acknowledged that a large degree of concurrency was planned but stated that "...concurrency is unavoidable in a program with many of the system components still in development..." and also that, "...due to the long production time required for the MCLWG, it is not feasible to have the related development project completed prior to commencement of MCLWG production." Concurrency risks were judged to be manageable. The DCP recommended option 1 (of five options presented, including termination) which called for commencing procurement in FY79 and procuring 32 MCLWGs.

T&E ASSESSMENT: The T&E assessment dated 4 November 1977 agreed that the gun and mount had performed well in tests but since many elements had been missing, stated that additional testing was required to integrate the total system and demonstrate overall system operational capability. It concluded that "... MCLWG will be operationally effective if program objectives are met for improved ammunition and other systems now under development." Consequently, the DDT&E recommended that the Navy develop an integrated program structure to ensure early testing of essential systems prior to the initial buy of five guns and that a new risk assessment of the other systems under development be prepared.

DSARC DECISION: The milestone III review was held on 8 November 1977. The resulting memoranda from the USDR&E (2 December 1977) and Deputy Secretary of Defense (6 January 1978) authorized limited production of the first seven guns of a 17 gun procurement (option 1 of the DCP). The recommendation for limited rather than full production was based on the fact that there were "... concurrent ammunition and related ship system component developments during the 4 years required for production of the first gun...." In authorizing the limited production, the Deputy Secretary of Defense requested that a program review be held in mid FY80 before release to full production "...in time to effect the FY81 procurement...."

EXHIBIT V-2
Historical Schedule



November 1977, the program was cancelled shortly thereafter. The program did not have a MENS, but there was a Navy operational requirement (OR). No DSARC II was held.

The question of inadequate T&E data primarily concerned the fact that the MCLWG was being tested on a World War II destoyer rather than on the new DD963 on which it would have been installed. This meant that the gun had to be tested with an obsolescent analog fire control system (MK62) instead of the digital MK86 for which it was designed. In addition, the next generation improved conventional munitions (ICM) and semiactive laser guided projectiles (SALGP) which would have given the gun enhanced capability were not yet available.

While a TEMP did exist before the major DT&E and IOT&E (although just barely), it was not approved by OSD or even forwarded there for comment. This action was not required by existing policy nor was the program even designated as a major system at the time. OSD involvement grew out of a controversy between the Naval Sea Systems Command (NAVSEA) project personnel and the independent tester, Operational Test and Evaluation Force (OPTEVFOR).

DT was conducted on land-based (1971) and shipboard (1975) facilities and demonstrated that the system performed functionally and that contract technical requirements had been met. Shipboard test used the DD-945 as a surrogate for the DD-963. R&M was poor initially, but improvements were made. OT III operational evaluation (OPEVAL) was conducted from September 1975 to June 1976 and verified the DT results in an operational environment. However, the tests were limited in scope due to the lack of the DD-963, the new ammunition, and associated systems.

In order to settle some earlier disagreements, the Chief of Naval Operations (CNO) in a letter dated 1 November 1973 set forth the acceptance criteria for operational effectiveness and suitability of the MCLWG. Nevertheless, the Commander, OPTEVFOR (COMOPTEVFOR) concluded that the MCLWG would not be operationally effective based on the mission requirements stated in the SOR (not including improved accuracy and effectiveness that were planned with the SALGP and ICM ammunition). Only after CNO intervention were approval for service use (ASU) and plans made for initial production. However, concurrent GAO studies tended to support the OPTEVFOR results showing a lack of accuracy and overall effectiveness. GAO recommended that OSD review the program before proceeding with production.

At the same time, Congress, responding to the perceived value of large-caliber ammunitions during Vietnam and the decommissioning of the Navy's last 8-inch cruiser, was exerting extraordinary pressure to proceed with, or even accelerate, the program. All of this attention caused the program to be elevated to formal DSARC review status. OSD insisted that the MCLWG be viewed as a system to include the MK86 fire control, the new ammunition, and other systems with which it would eventually be used. They directed that a DCP and integrated program plans be prepared and staffed and that all of the other requirements for a review of a major system be met. In the meantime, OSD deferred the FY77 production money (already appropriated by Congress) pending the outcome of the DSARC.

In light of the differences in maturity of the MCLWG and its associated systems, the DSARC decision to proceed with limited production and to require a follow-on OT&E and DSARC IIIB review of the performance of the integrated system would appear reasonable. What eventually led to the program's demise was the disconnect between the DSARC and the PPBS.

Shortly after the DSARC, the Navy, while constructing POM 80, realized that there was insufficient out-year money to procure guns at the rate just approved by the DSARC. Because of OSD's attitude to force a choice between alternatives, the program dropped out of the Navy POM and, without the out-year money, the FY78 initial procurement was also cut. Eventually, the entire program was cancelled.

In the final analysis, the MCLWG program was less affected by inadequate T&E data than it was by the lack of a consensus regarding the need for the program and the lack of harmony between the DSARC and PPBS. Nevertheless, the following observations are warranted:

- From a T&E standpoint, the decision not to allocate the resources associated with the DD-963 was key. As one former project officer said, "We got an obsolete ship, a primitive fire control system, and ancient ammo."

- The difference in maturity between the MCLWG and its associated systems was clearly recognized by the DSARC. Given the long production lead time of the gun, the limited production decision was probably a rational response to the inadequate data available.

- The controversy regarding the results of OTII is not necessarily an indication of a breakdown in the T&E process. Although strict adversarial relationships are seldom beneficial, checks and balances are a desirable feature that often prevent large-scale mistakes. Some in the development community felt that OPTEVFOR had exceeded its authority and adopted an attitude of recommending cancellation of any system that failed to meet its specifications. However, it may be that the misunderstanding derived from a difference of opinion on the scope and objectives of the OPEVAL tests.

- The controversy surrounding milestone III and the eventual cancellation of the program also could be attributed to the absence of a milestone II decision that would have affirmed the mission need and established meaningful thresholds and constraints.

2. AEGIS WEAPON SYSTEM

The Aegis DSARC IIA was held in June 1974 and directed the Navy to "... effect simplification of the AEGIS design and to demonstrate Aegis/Standard Missile (SM-2) compatibility." Three and one-half years later, in January 1978, OSD approved production of the system in spite of substantial evidence of inadequate T&E data. A summary of the formal documentation surrounding the DSARC is shown in Exhibit V-3; a historical schedule is shown in Exhibit V-4.

Shipbuilding is a special case in the acquisition process. It has more in common with the construction industry than it does with the development and production of most other types of weapon systems. The long lead times, large capital investment, complexity of integration with associated systems, and the relatively small number of units produced (compared to missiles or even airplanes) blurs the distinction between R&D and production and makes the term "limited production" meaningless.

Aegis is no exception. Considering that the Aegis weapon system and the CG-47 shipbuilding program were merged for the milestone III review, it was an enormously complex program whose total cost was over $15 billion. One would expect that a program of this size would be controversial. In fact, it was during its early development that the Navy was having internal arguments about which platform to put Aegis on. However, after the sinking of the Israeli Flat by a Russian-made antiship missile in [illegible] with consensus on the urgent need for an

EXHIBIT V-3
Decision Documentation
Aegis Weapon System

TEMP: TEMP #124 was published in April 1976 and approved within the Navy by OP03. It was sent to DDT&E but there was no formal review or approval. A revised TEMP #124(Rev. 1) was prepared prior to DSARC III.

PRE-DSARC III DCP: DCP #16 (Rev 2), dated 18 November 1977, concluded that "...technical risk for the AEGIS Weapon System is low," that "...extensive Development Test and Evaluation (DT&E) and Operational Test and Evaluation (OT&E) of Aegis in USS NORTON SOUND have demonstrated that the system has no residual risks," and that the integration risks are related to "...computer program completion." The DCP concluded that "there were no T&E issues which offer high risk," and recommended: "Proceed with the Aegis Weapon System Acquisition Plan commencing in FY78 and structured to support the Aegis shipbuilding program."

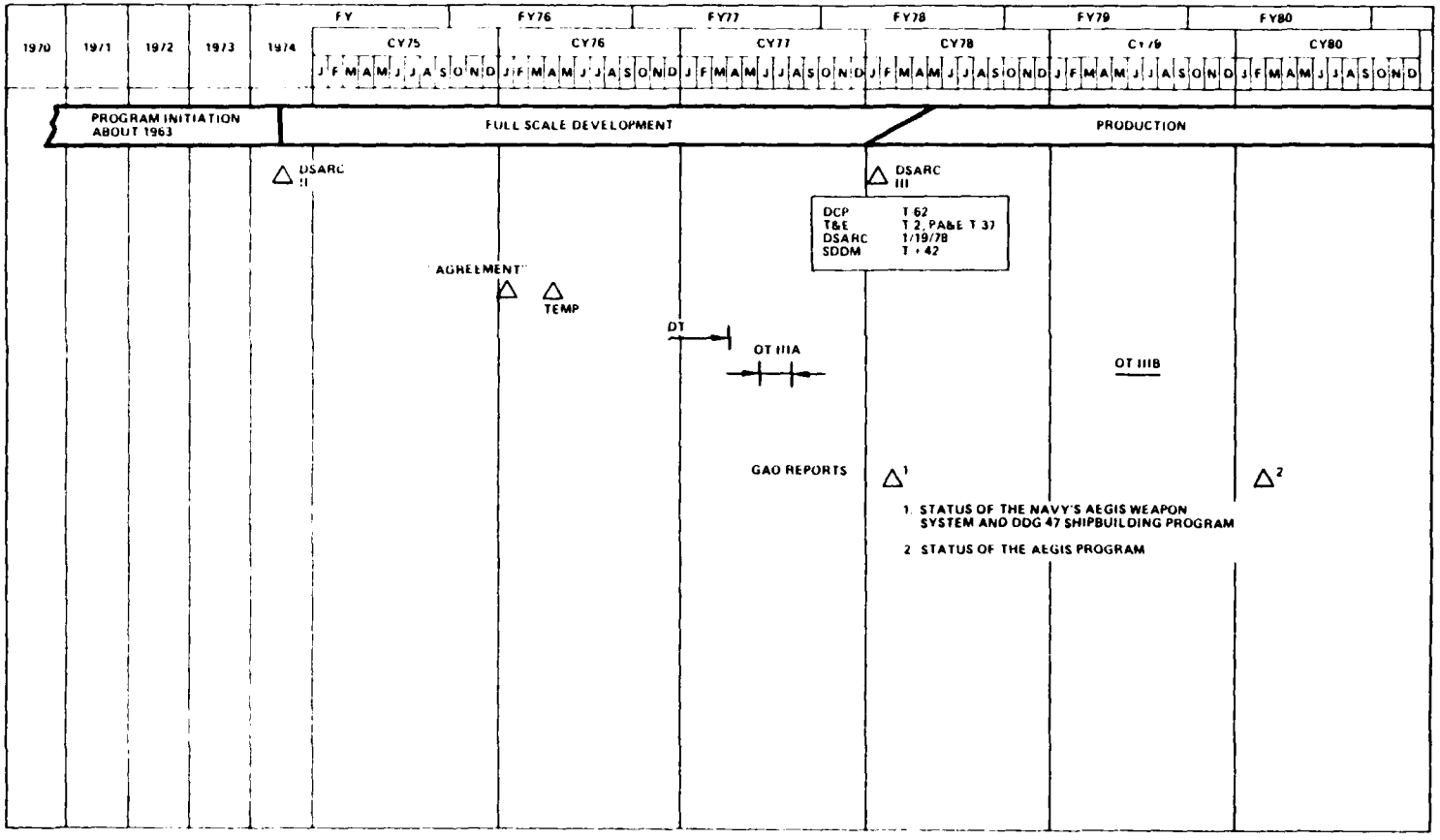
T&E ASSESSMENT: The T&E assessment dated 17 January 1978 stated that Aegis had completed considerable DT&E and IOT&E and that OT IIIA had demonstrated potential over existing systems in spite of the absence of key software, C^2, and ORTS elements. Software requirements were judged to pose moderate technical but high schedule risks. The DDT&E also concluded that although "...OT to date is inadequate to assess the effectiveness and suitability of the Aegis Weapon System in a realistic threat environment," that follow-on testing would add confidence. He recommended that production of one Aegis and installation on the DDG-47 were warranted. Follow-on production would be subject to the results of additional DT and OT of EDM-3 at CSEDS in Morestown, N.J.

PA&E ASSESSMENT: In a memo dated 13 December 1977, the ASD(PA&E) raised two issues for DSARC consideration: (1) use of the Army's Patroit missile instead of the SM-2, and (2) effective operation of the system in a high-threat environment.

DSARC DECISION: The milestone III review was held on 19 January 1978. The resulting Deputy Secretary of Defense memo, signed 2 March 1978, authorized the Navy to proceed with production of the Aegis and installation on the DDG-47 as recommended subject only to submission of specified planning and cost data to OSD.

EXHIBIT V-4
Historical Schedule

AEGIS



improved antiair warfare capability for U.S. Naval surface forces. Consequently, Aegis enjoyed a priority within the Navy second only to TRIDENT.

Aegis was begun under the total package procurement (TPP) concept and converted to the DSARC/milestone review system shortly after its inception in 1969. Planning for the DT/OT tests in support of the milestone III review centered around an agreement in 1976 between the DDT&E and the Navy regarding sequential test phases designated IIIA, IIIB, etc., and incremental DSARC reviews. It was at this point the DDT&E acknowledged that, subject to the concurrence of COMOPTEVFOR, the successful completion of IOT&E on the USS NORTON SOUND would provide sufficient justification for release of the lead DDG-47 and first production of the Aegis system.

Project personnel, OPTEVFOR, and the technical community all participated in drafting the TEMP which implemented this agreement. It covered in detail the objectives of OTIIIA and the high-level parameters that would be demonstrated. It called for operation and maintenance by service personnel when appropriate and it made reasonable tradeoffs with regard to realistic threat scenarios. The TEMP was then viewed as the controlling planning document on T&E matters. "It was the Bible," said one individual.

And, according to some, it was a model TEMP. It was one of the first to be prepared in accordance with the May 1975 revision of DODI 5000.3 (and the subsequent revision to OPNAVINST 3960.10), and it was held as an example for other weapon systems.

Testing of the Aegis system was conducted on an engineering development model (EDM-1), a single quadrant early version of the SPY-1 radar, installed in the USS NORTON SOUND. Through December 1976, 28 missile (SM-1) firings were conducted by Navy crews; 21 were successful, 7 were attributed to missile failures. DT IIIA of EDM-1 SM-2 was concluded in May 1977 and consisted of nine SM-2 at-sea firings (six successes and one partial success). OT IIIA was conducted between June and July 1977 with 20 days at sea at the Pacific Missile Range. The system was evaluated in numerous nonfiring exercises, and there were seven SM-2 firings during which the system was operated and maintained by the USS NORTON SOUND crew, except when specific external maintenance assurance was authorized and documented.

The DT and OTIIIA tests were mostly successful and both COMOPTEVFOR and DDT&E concluded that the substance of the 1976 agreement had been met. Residual concerns were

that, in reality, only a small portion of the entire Aegis weapon system had been tested and that numerous questions remained unanswered. Missing pieces included full coverage from the SPY-1 radar, key software, elements of the command and control system, and the operational readiness tests system (ORTS). In addition, DDT&E observed that in spite of the progress, testing to date was inadequate to assess the system in a realistic threat environment.

In light of all these inadequacies, it may be surprising that the DSARC recommended that Aegis proceed into full production. There are several explanations:

- Project personnel argued that the tests conducted demonstrated successful operation of the high risk elements of the system; that the body of data was consistent with the plan to move incrementally to the next phase.

- Early agreement with OSD was key. The ground rules for proceeding into production had been negotiated well in advance.

- The program had a high national priority. Furthermore, there had been years of consensus building within the Department of Defense via interagency briefings and reviews. One person remarked that it was inconceivable that the DSARC would do anything other than recommend full production.

- The program had maintained relatively low profile; i.e., there had not been major overruns or other adverse publicity. The GAO was only routinely critical and there was strong Congressional support.

- The emphasis on demonstrated operational suitability had not yet taken hold.

- The nature of ship acquisition made full production essentially the same as limited production, which might have been more likely for other kinds of weapon systems given the unknowns that still existed.

Interestingly, the major opposition to production of the Aegis came from the ASD(PA&E) who was given specific responsibility within OSD for OT&E shortly after this DSARC. PA&E was concerned that there was insufficient

evidence that Aegis would function effectively in a high-threat environment. He also wanted to know why the Navy could not use the Army Patriot missile rather than the Navy SM-2.

3. EF-111A TACTICAL JAMMING SYSTEM

The EF-111A DSARC II met in October 1974. The resulting Deputy Secretary of Defense memorandum of 23 January 1975 approved a 38-month FSD program (instead of the faster paced 28-month program proposed by the Air Force). It is also stated that there were important questions regarding cost and operational feasibility that could only be answered by testing and that the Air Force should "... complete the test program set forth in the DCP with satisfactory results prior to the DSARC III review." An updated DCP (No. 126B) was issued on 13 May 1975. A summary of the decision documentation surrounding the milestone III review is shown in Exhibit V-5; a historical schedule is shown in Exhibit V-6.

The EF-IIIA milestone III decision is a study in contrasts. The Air Force was convinced that, while they were forced to cut back on testing due to overruns in R&D, the DT/OT they did conduct was adequate, that all performance thresholds stated in the DCP were met, and that their contract with DDT&E (via the TEMP) had been satisfied. On the other hand, DDT&E and the GAO were just as convinced that development test and evaluation were incomplete and that operational testing was too limited to provide a realistic threat environment for evaluating operational effectiveness and suitability. Specifically, they noted that there were 209 technical/design deficiencies reported by the Air Force Test and Evaluation Center (AFTEC) of which perhaps 40-50 were duplications. The majority of these were in the operability, reliability, and maintainability area. In addition, the GAO in particular remained skeptical of the EF-111A's military worth or cost-effectiveness.

From a historical perspective it appears that OSD, in fact, did break their contract with the Air Force. Reliability and maintainability began to emerge as important themes in mid-1978, after the TEMP had been reviewed and approved. By this time, the ASD (MRA&L) and his staff were emphasizing R&M; some believe that they dominated the DSARC decisionmaking process during this period. Their concern regarding the lack of "blue suit" maintenance may have been justified, but the intention to use contractor maintenance for IOT&E was clearly specified in the TEMP. The indications are that this milestone decision was at the leading edge of new policy initiatives and that it became the test case for those in OSD that wanted to set

EXHIBIT V-5
Decision Documentation
EF-111A Tactical Jamming System

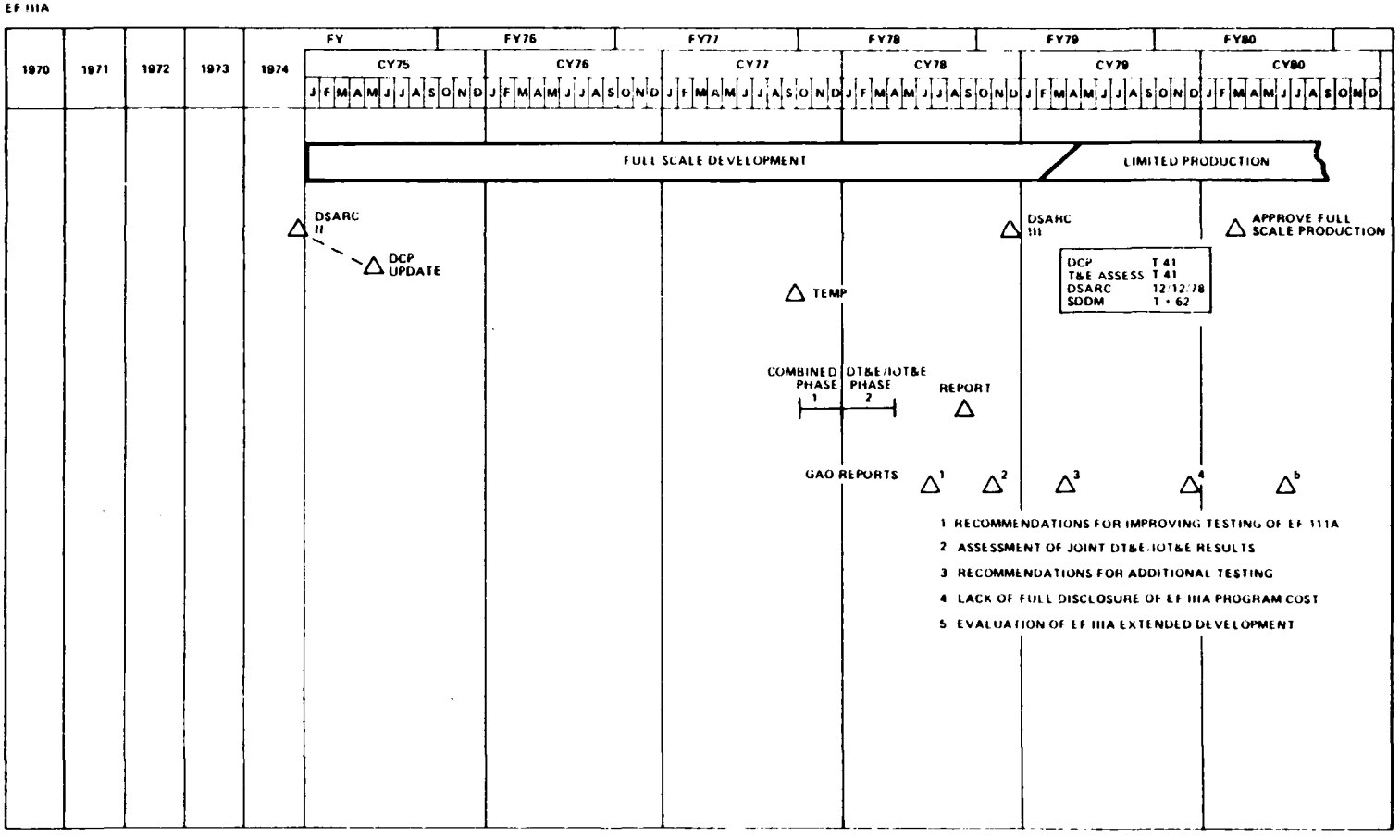
TEMP: A revised TEMP covering plans for the combined DT&E/IOT&E was published on 15 September 1977. It was reviewed and approved by DDT&E prior to the start of testing, although not early enough to significantly alter test resources or plans.

PRE-DSARC III DCP: DCP #126C (draft) issued 1 November 1978 stated that "...all DT&E/IOT&E planned objectives were met during the Phase 1B combined test program. It is the consensus of these organizations [SPO, TAC, and AFTEC] that the testing was as comprehensive as existing U.S. test resources would support and that additional testing is not required..." It also stated that all performance thresholds of the 13 May 1975 DCP were met and requested authority to proceed with rate production of 42 aircraft and parallel correction and demonstration of deficiencies.

T&E ASSESSMENT: The T&E assessment dated 1 November 1978 stated that there was a lack of adequate test assets and facilities to assess performance under certain conditions; that there were overall and specific maintainability and reliability problems, including electromagnetic interference (EMI), software, inadequate support equipment (and others), and limited evidence of "blue suit" maintenance participation; and that "...a favorable production decision cannot be supported until (these) operational suitability problems have been corrected and postively demonstrated..."

DSARC DECISION: The milestone III review was held on 12 December 1978. The resulting Deputy Secretary of Defense memo dated 12 February 1979 concluded that IOT&E was generally favorable, but the "...assessment of the operational suitability results disclosed a number of problems in reliability and maintainability..." It directed a phased production program with a parallel suitability evaluation "...to serve as a precaution against future increases in cost which might arise from the need to redesign and retrofit unreliable equipment..." The Air Force was directed to provide reports for OSD review at 3-month intervals with a recommendation for program continuation or termination. Based on these reports, the Acquisition Executive would determine the advisability of full-scale production.

EXHIBIT V-6
Historical Schedule



an example. Furthermore, the EF-111A was the first major program to have a DSARC III in nearly a year. Some believe this gave OSD staffers extra time to investigate and prosecute their emerging beliefs--that, in effect, they were ready and waiting for the EF-111A review. One OSD official said, "There were new cooks and a clean kitchen. When the meat was put on the table, everyone dug in..."

It appears that the GAO also adopted the EF-111A as a special case. They had just completed several reports on operational testing and, in particular, reported in December 1977 that "Air Force DT&E needs improvement." From June 1978 to August 1980, there were five separate GAO reports on the EF-111A--roughly one every 6 months.

The decision to limit production and conduct additional T&E was more a product of the times than of the previously established way of doing business. Subsequent policy changes have emphasized R&M issues and implemented procedural changes that would have facilitated the EF-111A milestone review. These changes notwithstanding, the limited production decision was a controversial one. Project personnel argued that it unnecessarily added 2 years to the program and increased total program cost by $100-$200 million. In addition, some feel that stretching out the production from 3 to 4 years was motivated by fiscal constraints rather than by a risk assessment of technical problems.

Adequacy of T&E data was clearly an issue during the decisionmaking process. It was highlighted in the DCP, primarily as a result of the GAO claim that "... accomplished testing was inadequate to proceed with a DSARC III decision." Other observations on events leading to the review include:

- R&D problems, including a $15-million overrun, were the primary cause of the reduction in the number of test aircraft (from two to one), the combined DT/OT, and the reliance on contractor maintenance.

- The test program was conducted essentially as specified in the TEMP; however, the program was overtaken by changes in policy that altered the expectations of the DSARC principals and left the Air Force largely unprepared for the review process.

- The Air Force felt the deficiencies found in IOT&E could be corrected by engineering change

proposals (ECPs) during full-scale production. OSD insisted on a more cautious fix-and-retest approach.

- The DDT&E found that a lack of adequate test assets and test facilities prevented an assessment of the EF-111A's ability to operate in certain environments. GAO agreed and recommended that the DOD improve its capability for testing electronic warfare systems and undertake a dedicated effectiveness test when improved threat radar simulators were in place.

- The DSARC was informed of the cost of delaying the EF-111A program in order to conduct additional T&E, but did not share the Air Force's sense of urgency for the system. It opted for an approach that would reduce the risk of subsequent redesign and retrofit.

4. ROLAND MISSILE SYSTEM

Full-scale development on Roland began as a result of a DSARC II held in February 1974. Subsequently, the Army evaluated three foreign systems that were under development as well as improvements to the U.S. Chaparral system. In January 1975 the Deputy Secretary of Defense approved the Army's source selection recommendation to acquire the Franco-German Roland II. Since that time, there were two major redirections in the programs: (1) as a result of cost growth, a special DSARC held in September 1976 restructured the program to reduce the number of missile firings from 146 to 100, and (2) again in response to cost growth, a special program review by the DSARC principals in June 1978 restructured the program, released FY78 IPF and other funds, and requested an increase of testing at the performance boundaries.

In spite of significant doubts as to the reliability and maintainability of the Roland missile system, limited production was approved in May 1979. The decision documentation is summarized in Exhibit V-7; a historical schedule is shown in Exhibit V-8.

The Roland program is the only one of the six selected for further study that had a heavy multinational flavor. In fact, the terms DT and OT were never formally used in deference to the phrase, technology transfer, which was supposed to reflect the maturity of the system and the commonality of U.S. Roland with its Franco-German counterpart. Undoubtedly, the North Atlantic Treaty Organization (NATO) influence added another dimension to the milestone review and further complicated the question of T&E adequacy.

EXHIBIT V-7
Decision Documentation
Short Range Air Defense (SHORAD) System
US ROLAND

TEMP: A formal TEMP was not available during the FSD phase, although a draft TEMP was submitted just prior to DSARC III in April 1979. The US Roland test program was based on the test philosophy contained in the April 1976 "Powers Report" (see text).

PRE DSARC III DCP: DCP #95, Rev. C dated 22 March 1979, "For Comment Draft" addressed the issue of whether Roland was ready to enter the production phase. The DCP did not include a specific recommendation although six program alternatives (from termination to acceleration) were described.

T&E ASSESSMENT: In his memorandum dated 30 May 1979, the DDT&E concluded that testing "...had verified that European technology has been successfully transferred and that Roland has essentially met..." the system performance parameters contained in the Army ROC. In addition:

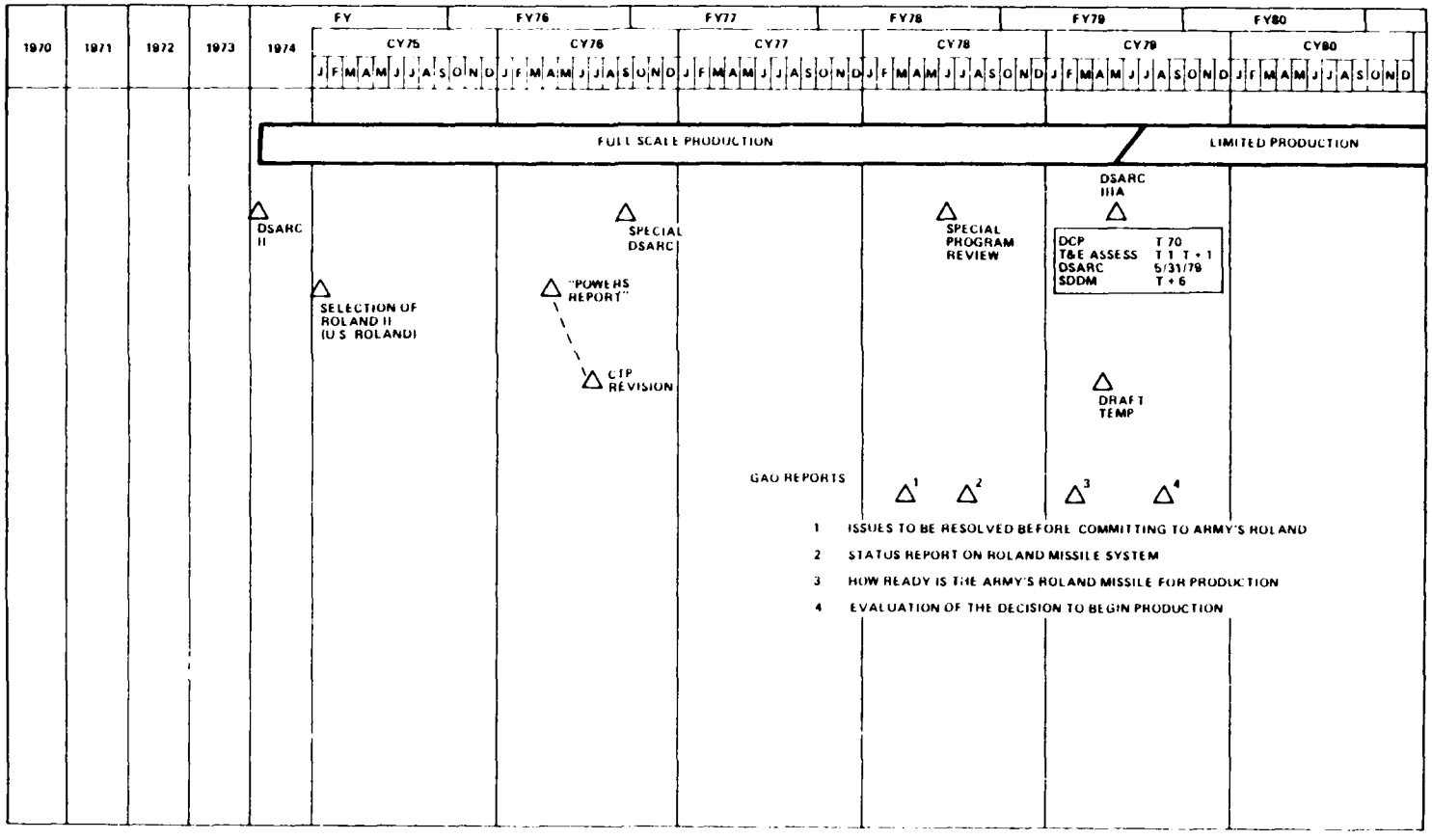
- That a significant effort would be required to achieve the RAM goals
- That the system exhibited susceptibility to downlink jamming and additional testing was required to verify performance
- That adverse weather testing conducted was inadequate
- That maintainability was not tested.

Consequently, the DDT&E recommended that any production beyond low-rate production should be contingent on demonstrated RAM growth, and that modified test plans and results of planned tests be forwarded to OSD for review. In a subsequent memo dated 1 June 1979, the DDT&E further stated: "Testing in the areas of ECM susceptibility, maintainability, adverse weather performance and warhead lethality was either incomplete or has yet to be accomplished."

DSARC DECISION: The DSARC convened on 31 May 1979 and determined that the Roland program should be pursued and that the Army should be authorized to enter limited rate production. In his memorandum to the SecDef/DepSecDef dated 4 June 1979, the USDRE also stated: "When the results of currently planned reliability tests are available, we will conduct a DSARC IIIB review to determine the appropriate time to enter full-scale production." The SecDef decision memorandum dated 6 June 1979 stated that "Roland acquisition cost estimates, current progress, and performance are satisfactory for limited production," and granted approval for low rate initial production (LRIP) in FY79 and FY80. It also directed the Army to provide information on proposed support concepts to ASD (MRA&L) within 120 days and a revised TEMP to DDT&E within 45 days. Finally, it called for a DSARC IIIB review of RAM capability before entry into full production.

EXHIBIT V-8
Key Events

ROLAND



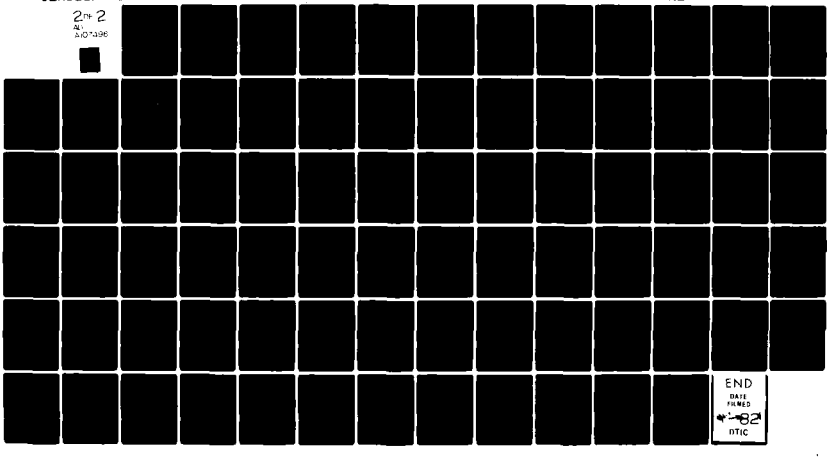
A TEMP was not prepared or submitted for OSD review until just prior to the DSARC III because, according to PM personnel, "TEMPs weren't required until early 1979." (Actually, the revision of DOD Directive 5000.3 that required submission of a TEMP to OSD for review was issued in April 1978.) Fortunately, there was very early coordination on an equivalent document, the Army coordinated test plan (CTP). One member of the PMO said, "DDT&E was in bed with us since 1976 and the OSD staffers were active participants in our Test Integration Working Group (TIWG)."

This high degree of coordination stemmed from an earlier review of Roland testing, which was headed by MG Powers, then Commander of the Test and Evaluation Command (TECOM), and subsequently referred to as the "Powers Report." In 1975 the Roland program was subject to severe funding purturbations which upset the existing T&E plans. The purpose of the Powers report was to restructure the test program in light of the reduced funding and to effect efficiencies in the test process. It essentially laid the ground rules for T&E through milestone III. Even though it substantially reduced the amount of testing to be conducted, it acted as an implied contract between OSD and the Service. At the time of adoption, the schedule for the test program was acknowledged to be "... optimistic and of high risk, but the data produced will permit a moderate to low risk production decision." Further, the report noted "... a major deviation from the standard acquisition process in that the production decision will have been made without an adequate evaluation of RAM, training, and logistics support." As one PM official put it, "We knew this was the least we were going to get away with..."

The flight test program consisted of 107 missile firings of which 70 were without failures, 31 had failures of some sort, and 6 were declared no tests. The flight test program identified quality control and design deficiencies which, according to the DCP, were subsequently resolved. In addition, a nonflight test program was conducted which identified design shortfalls. These included electronic countermeasures, reliability, and human factors. The independent evaluation of the operational effectiveness and suitability of US Roland conducted by the U.S. Army Operational Test and Evaluation Agency (OTEA) concluded that the system had demonstrated capability to perform the all-weather mission although reliability was insufficient and maintainability had not been tested. OTEA concluded that system effectiveness and reliability should be further examined in follow-on evaluation.

However, the program was subject to other forces during its development. One of those came with the change in administrations in January 1977 when Dr. Perry replaced

AD-A107 496 BOOZ-ALLEN AND HAMILTON INC BETHESDA MD F/G 5/1
THE RELATIONSHIP BETWEEN ADEQUACY OF TEST AND EVALUATION DATA A--ETC(U)
JUN 81 MDA903-80-C-0591
UNCLASSIFIED NL



Dr. Currie as the USDRE. Previously, the management philosophy, at least as perceived by the Project Management Office (PMO), was, "We know you have problems, but get it in the field and we'll fix it later." Later, that changed to "Fix it and prove you've fixed it before we talk about production." In addition, there was pressure from the GAO and Congress to provide satisfactory evidence of Roland's performance before committing to production. Some even wanted to revisit the whole selection of Roland as the Army's future all-weather short-range air defense system.

OTEA and DDT&E agreed that the system's shortfalls included R&M and ECM and that adverse weather and maintainability testing was inadequate. GAO raised similar objections but DOD responded by saying that the GAO report was particularly misleading about the adequacy and amount of testing to be completed prior to the production decision.

The R&M issue was perhaps the most serious. The Roland DSARC followed by less than 30 days the DSARC III on the XM-1 tank which had serious reliability and maintainability problems. This criticism was also levied at Roland. It appears that Roland became another victim of the generally critical atmosphere that began with the EF-111A (in December 1978). The DDT&E had concluded that failure to demonstrate reliability goals was a chronic problem and that it required a change in the way DOD was doing business. One official said "Once the gun had been cocked and fired, it was easier to use."

Cost was another issue which may have impacted the DSARC decision to limit production. The priority of air defense had softened somewhat since the Roland program was started and Patriot was consuming the majority of what was left in the FY80 budget and the FY81 POM. One theory is that out-year cost increases due to uneconomic production buys were accepted in order to meet current year dollar constraints; i.e., that the T&E issues were a convenient way of addressing the budgetary pressures associated with the "bow wave."

Finally, while the Army may have been willing to trade Rolands for Patriots, the proponents of RSI and the "two-way street" observed that Roland was the only foreign system being considered for production by the United States. One of the specific issues identified for the DSARC was, "What are the RSI implications of not procuring Roland?"

The Roland program may or may not be typical, but it provides an interesting vignette: (1) test planning was coordinated well in advance; but (2) cost and schedule

pressures forced the Army to conduct limited IOT&E on P&D hardware not representative of the production configuration; further, (3) new philosophies regarding decision criteria emerged after the original plan; this resulted in (4) inadequate T&E data to support a production decision; however, (5) OSD decided to go ahead with limited production due to other factors.

5. AIR-LAUNCHED CRUISE MISSILE (ALCM)

The ALCM program was initiated in July 1973, and a DSARC I was held in February 1974 which approved entry into an 18-month advanced development phase. In December 1974, a DSARC II review resulted in disapproval of entry in the full-scale development phase. A subsequent DSARC IA review in March 1975 resulted in approval of an advanced development demonstration for proof-of-concept. During the period FY75 through the first quarter of FY77, the Air Force and the prime contractor, Boeing, conducted tests intended to demonstrate system integration in a flying vehicle within established threshold performance parameters. Subsequently, a DSARC II review was held in January 1977 which approved the Boeing design (AGM-86) for FSD. The DSARC also directed the establishment of a Joint Cruise Missile Project Office. Nine months later, in September 1977, the USDRE reoriented the program by directing a competitive development between the Boeing AGM-86 and the General Dynamics AGM-109 (an air-launched version of their BGM-109 sea-launched cruise missile). In February 1980, the Air Force selected the AGM-86B configuration for production and deployment.

The air-launched cruise missile program became the major moderization effort for the strategic bomber force when the B-1 was cancelled by President Carter in mid-1977. The new administration's Under Secretary of Defense for Research and Engineering, Dr. William Perry, was a strong proponent of the technology used in ALCM and other members of the cruise missile family, and he provided a focal point for support of the program within OSD. Perhaps because of this support and because ALCM was a critical part of the strategic force planning, it was approved for limited production in April 1980 in spite of a large number of test issues which were inconclusive or undetermined. The DSARC documentation is shown in Exhibit V-9 and the historical schedule in Exhibit V-10.

ALCM was an ambitious, success-oriented program with a high degree of planned concurrency. IOT&E was to be conducted on immature and incomplete systems. For example, integration into the B-52 aircraft was to be accomplished after the production decision. Combined DT&E/IOT&E tests were conducted with developmental prototypes procured from competing contractors.

EXHIBIT V-9
Decision Documentation
Air Launched Cruise Missile (ALCM)

TEMP: The 28 March 1979 ALCM TEMP was the baseline document for combined DT&E/IOT&E although tests were already underway by the time of approval. A draft revised TEMP was provided to DDT&E in anticipation of the DSARC III review.

PRE DSARC III DCP/IPS: DCP #136 dated 3 March 1980 supported a milestone III decision by recommending "...that the ALCM be approved for full production as outlined in this DCP with the provision that the FOT&E program be closely monitored to assure that performance and reliability thresholds are achieved." Because the ALCM Source Selection was not completed at the time, the DCP included demonstrated performance parameters but did not discuss specific results of tests obtained during the competition. Threshold achievement and other data were to be reported at or before DSARC III. According to the DCP, the major issue was the risk associated with the reliability/maintainability of the system. The DCP also stated that "...follow-on development and operational test and evaluation will adequately address remaining issues of risk and concern."

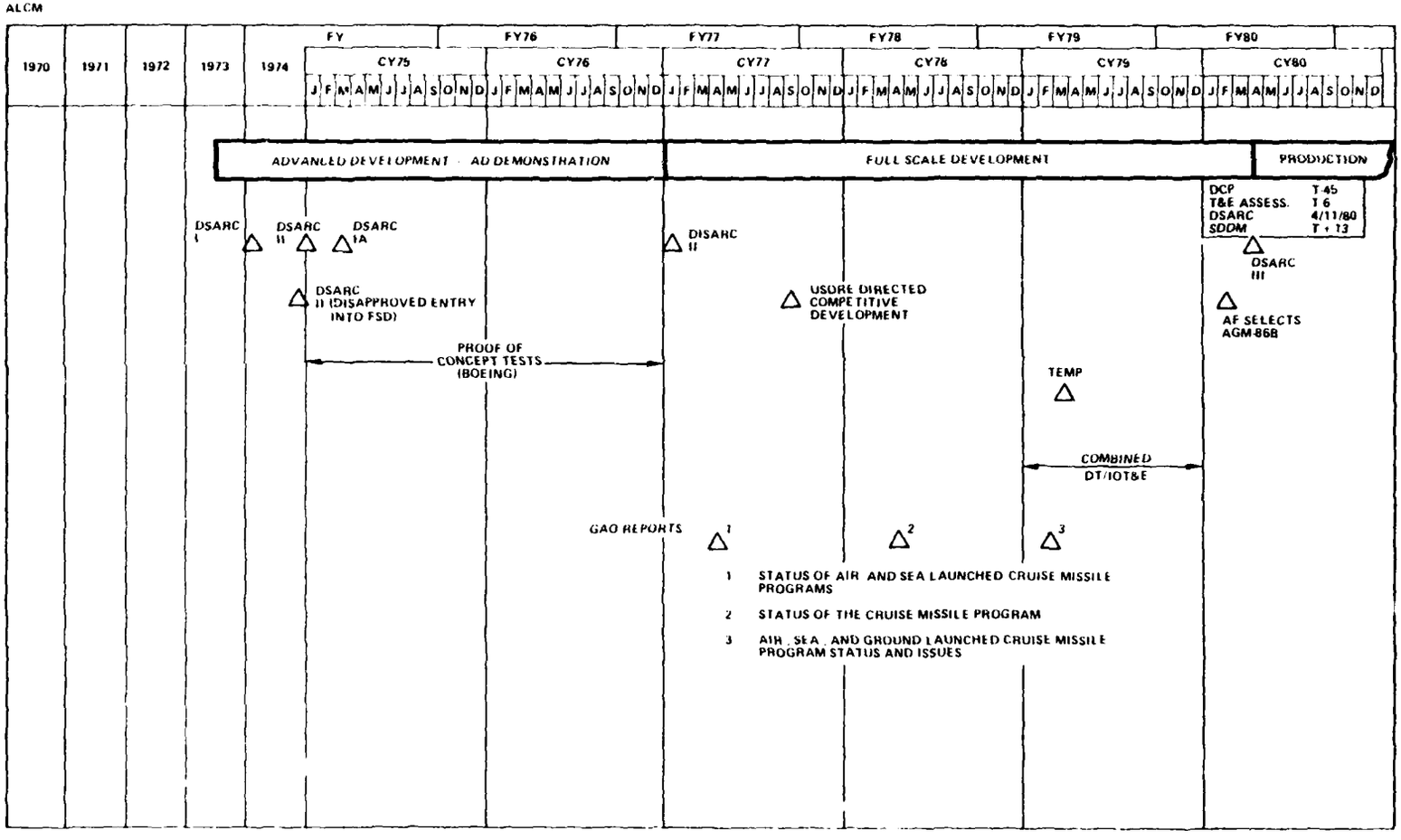
T&E ASSESSMENT: The DDT&E submitted his independent T&E assessment in a memorandum dated 11 April 1980. He stated that significant data were obtained on most of the key performance parameters but that some desired tests were "... impacted by the limited number of flights (four were shortened by malfunctions) and somewhat hindered by instrumentation problems." The DDT&E reported that a substantial amount of DT and OT remained to be done because of subsystems still undergoing development (radar altimeter, B-52 OAS, and support equipment). Although the Air Force's proposed combined DT&E/FOT&E emphasized these areas, the DDT&E concluded that the DT and OT test results to date indicated that the "...ALCM air vehicle is potentially capable of meeting operational requirements." However, he also cautioned that "...more extensive and realistic operational testing with production representative missiles, carrier aircraft, and support equipment is essential to successful production and deployment of ALCM." Finally, he urged timely completion of the remaining development and early operational testing of those systems. He suggested subsequent, periodic OSD reviews of follow-on test results.

EXHIBIT V-9 (continued)

DSARC DECISION: The DSARC III was held on 17 April 1980. The resulting SDDM, dated 30 April 1980, approved production leading to a minimum capacity to produce 40 missiles per month, established reliability and maintainability thresholds and goals, and directed the Air Force to take certain actions as recommended by the DSARC. The significant T&E related directives included:

- Formal submission of an updated TEMP witnin 60 days that defines a program for determining key ALCM parameters for which limited test data exist and for evaluating systems still under development
- Provision of reports to OSD in October 1980 and March 1981 prepared by AFTEC concerning progress made in the test program
- Continuation of RAM efforts and report to the Cruise Missile Executive Committee.

EXHIBIT V-10
Historical Schedule



The test program involved ground testing of the integrated air vehicles, including evaluation of radar cross section, infrared, electromagnetic pulse (EMP), (IR) signature, (EMI/electromagnetic compatibility (EMC), engine/air frame compatibility, engine performance, aerodynamic performance (wind tunnel tests), warhead/impact fuze tests (sled tests), and maintainability). The tests were conducted on ten flights each of the AGM-86B and the AGM-109. The Air Force Flight Test Center (AFFTC) was responsible for the DT portion of the tests and AFTEC was responsible for the OT portion. Each test flight was designed to accomplish multiple DT and OT objectives. According to the DCP, "USAF personnel conducted 'hands-on' maintenance on the competing systems as part of the overall evaluation."

The pace of the test program--20 flights within an 18-month period--stressed both the developer and the tester. Furthermore, there were neither management reserves nor opportunities to recover gracefully from technical problems as indicated by test failures if they occurred.

In such a program, the TEMP becomes a critical document for communicating the ground rules to all concerned and ensuring there are no T&E surprises at the milestone review. It appears that the ALCM TEMP did just that. Once negotiated, it was viewed by the Air Force as a contract with OSD and, to the extent possible, it was strictly adhered to.

However, shortage of missiles, test failures, range limitation, competetive prototyping (which diluted the T&E effort), and the absence of subsystems still undergoing development resulted in a body of T&E data that was, in the end, judged inadequate to make a full production decision. AFTEC's independent evaluation cited eight performance parameters that were inconclusive or undetermined due to a need for more testing. These parameters included: overall operational effectiveness, terminal accuracy, terrain following, and launch envelope (for which definite estimates of performance could not be made because of the lack of test points and insufficient data); mission planning, software suitability, weapon system compatibility, and interoperability (because critical aspects of the system were not available for test). Three other parameters were judged deficient and 13 satisfactory.

The DDT&E concluded that more extensive and realistic testing was necessary. While there was some disagreement between AFTEC and the SPO on the interpretation of the IOT&E results, it was clear that there was still much uncertainty in the operational effectiveness and suitability of the system.

Nonetheless, the DSARC recommended that the Air Force proceed with limited production. In this case, it appears that the national priority and urgency for the program coupled with high-level DOD support tipped the balance in favor of accelerated development. Also, the decision to go ahead with the ALCM was made in conjunction with the decision to cancel the B-1 bomber, which was formally announced 3 months later. Furthermore, the reviewers knew in advance that the production decision would have to be made with less than full data and they accepted the associated risks.

6. PATRIOT MISSILE SYSTEM

The Patriot weapon system development program was initiated in 1965. Contractor tests and simulations were conducted from 1967 to 1972 to demonstrate the Patriot surveillance and guidance principles and evaluate hardware fabricated to initial design specifications. FSD began in March 1972 when the DepSecDef approved DCP #50. The development was redirected in 1974 when the DepSecDef instructed the Army to provide early flight verification of the track via missile (TVM) guidance principle while exercising fiscal restraint in other areas. A DSARC review was held in January 1976 in which the decision was made to resume full-scale engineering development. Since that time, prototype fire units were produced, significant design changes were made (particularly the conversion to digital modules in the missile guidance system) and a total of 50 contractor flight tests were conducted. As shown in the decision documentation, Exhibit V-11, the DSARC agreed to limited production of the system in August 1980 even though there was strong evidence of inadequate T&E data. A historical schedule is shown in Exhibit V-12.

Patriot gained a reputation as the longest development program in the history of the Army, if not the entire DOD. With the total development time approaching 20 years, there was enormous pressure on the PM and the Army to go to the DSARC as soon as possible in order to transition into the production phase and start putting something into the field. Ironically, this pressure, emphasized by the possible loss of funding or threats of cancellation, in itself contributed to the fact that inadequate T&E data were available to the decisionmaking process. As the milestone review neared, the Army decided to conduct OT prior to DT so that they would at least have partial DT and OT results to present to the DSARC. Unfortunately, this meant that things that should have been discovered and corrected in DT were not; and that, consequently, the system was not sufficiently mature to have a meaningful IOT&E. A senior OTEA representative said, "The tests were mostly negative, some positive and largely inconclusive."

EXHIBIT V-11
Decision Documentation
Patroit

TEMP: A CTP was prepared early in the FSD program and coordinated among the Army test community; however, there was never formal OSD approval. In preparation for DSARC III, the Army submitted the CTP dated August 1978 which was rejected as having too much detail for a management review. A TEMP submitted in April 1979 was rejected because it did not have enough detail regarding post-DSARC III plans. A revised TEMP, dated 11 July 1980, was subsequently approved.

PRE-DSARC III DCP: The Patroit Air Defense System DCP, dated 20 August 1980, described three alternatives for pursuing the production and deployment of the system (#1-remain in FSD, #2-accelerate the program, and #3-initiate a system confirmation program) plus one alternative to terminate the program. On 3 July 1980, the ASARC adopted alternative #2 as the preferred approach; however, the DCP states: "...to alleviate reliability and investment risks while maintaining an accelerated program, execution of the newly introduced alternative #3 is recommended. This alternative provides a prudent, yet quick, program to expand USAREUR air defense capabilities while reducing attendant risks and keeping program costs relatively low." In addition to the program alternatives, the DCP addressed four DSARC issues, including the impact of not testing certain elements of the system, the adequacy of existing test data, and the risks associated with the software development plan.* In each case, the DCP advances a positive case for the Patroit system. The DCP was supplemented by an Integrated Program Summary (IPS), dated 14 July 1980.

T&E ASSESSMENT: The DDT&E in his assessment dated 15 August 1980 concluded that "...the Patroit system, as tested, did not demonstrate its readiness for opera-

* Others included the number of ready missiles per fire unit, the performance of TVM in an ECM environment, and the electronic counter-countermeasures (ECCM) capability to meet the growing threat.

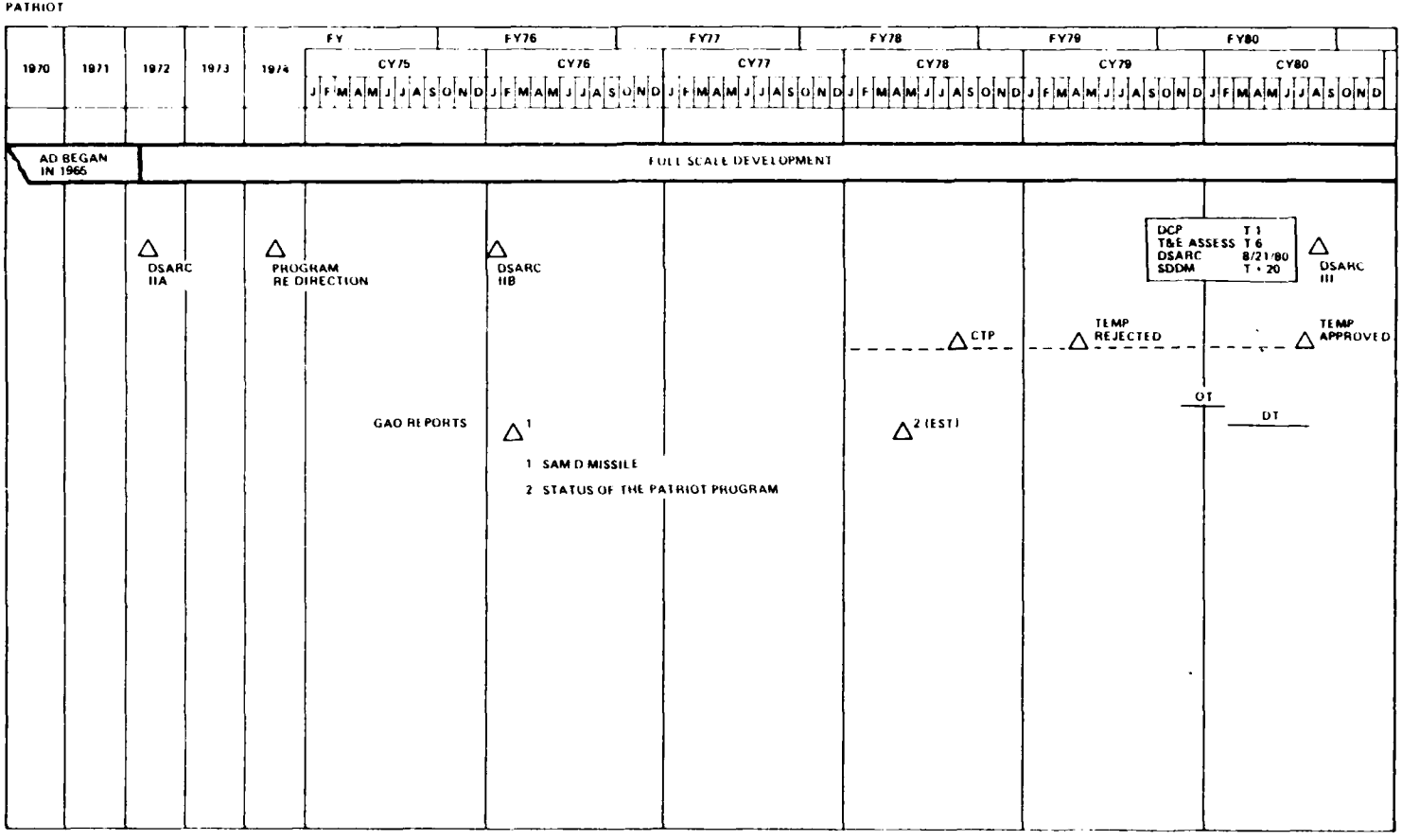
tional employment." Problems included missile reliability (particularly in an operational environment), inadequate ECCM under some conditions, poor fire unit reliability, unsatisfactory mobility, and immature and constantly changing software. In addition, because the system was maintained by the contractor during OT using nonstandard corrective maintenance techniques, the "...present data in fault localization and repair (were) of little use ..." and maintainability could not be determined. In the event of a decision to proceed at a minimum production rate (perhaps anticipating this outcome), the DDT&E recommended an extensive structured program of tests and reviews: "The test and review program should be established to insure that substantial progress toward achieving satisfactory Patroit operational suitability is demonstrated before production is permitted to rise above the minimum level."

DSARC DECISION: DSARC III was held on 21 August 1980. The resulting SDDM to the Secretary of the Army, dated 10 September 1980, took exception to the Army's request but approved "...limited production at a rate appropriate for confirmation of the design as manufactured on production tooling."

The Secretary of Defense's position was based on the fact that "...performance in ECM conditions has not been fully demonstrated by operational and development testing, nor has there been demonstration of acceptable levels of reliability and maintainability in either the fire units (FU) or in the missile." He further stated that "... hardware off production tooling in some quantity will be required in order to both prove the production design and to provide enough equipment for obtaining soldier experience, and that the limited production minimizes the financial risk incurred since anticipated changes will not result in appreciable retrofit cost."

In so stating, the SecDef established a comprehensive set of conditions that had to be met before release of subsequent increments of funding. He further required the submission of a revised TEMP, which followed the guidelines established in the DDT&E assessment of 15 August 1980, within 45 days and periodic reviews to determine the level of production justified by test results. He even directed that an award fee contract based on technical performance be used for the follow-on effort.

EXHIBIT V-12
Historical Schedule



The PM staff believes that if they had had another year, they would have had a smooth DSARC review, but that they could not afford to "get out line." That is, the lead times in the budget (PPBS) process are such that if they had not had the DSARC when they did and, in fact, gotten approval for some form of production they would have lost the FY81 money and perhaps risked termination of the the program. They also believe that the noted deficiencies in software, diagnostics, and R&M could have been fixed in parallel with a full production program. OSD disagreed.

From the PM's perspective, each of the OSD elements acted as an antogonist--pursuing narrow technical interests without attempting to integrate all the factors that impacted the cost of delay versus the cost of proceeding. However, it is also true that the contract between OSD and the Army regarding the adequacy of T&E data necessary to support a production decision never existed. Although the Army had prepared, staffed and was using a CTP, it was not until just before milestone III that a TEMP was approved by DDT&E. This was caused partly by the fact that Patriot was always one generation behind the management philosophy of the day; i.e., the program had been overtaken by so many changes in policies and procedures, not to mention personalities, that it was always in isolation some policy and, consequently, became fair game for any number of critics.

There were also indications of lack of cooperation within the Army. The PMO felt that TECOM dwelled on "... the 20 percent that went wrong rather than the 80 percent that went right," and that this bias carried over to the independent evaluators, the Army Materiel Systems Analysis Activity, AMSAA and OTEA. OTEA felt that much of the DT testing was not representative of an operational situation and refused to accept and use some of the data. They also felt that the system was fundamentally not ready for OT and it may be that they conducted the test under mild duress. This internal strife also existed at the Developmental and Readiness Command (DARCOM) and Army Headquarters was probably perceived by the OSD staff, each of whom had his own suggestions on how the program ought to be fixed.

In the end, however, the rising costs associated with continued development mitigated the inadequacy of the data and resulted in a limited production compromise.

7. SUMMARY OF FINDINGS

The six programs studied varied widely with respect to the specifics of their test and evaluation and the degree to which perceived inadequacy influenced the eventual DSARC recommendation. Each program and milestone review was very much a product of its special circumstances in the context of the existing decisionmaking environment. However, as shown in Exhibit V-13, some common factors are apparent:

- Three of the programs were extremely long, 11 or more years between the start of AD and the milestone III review. Long programs tend to be overcome by changes in acquisition and T&E policy, new management philosophies of key decisionmakers, and even the threat.

- Three of the programs had attracted a large amount of attention from the GAO and/or Congress and had achieved a degree of notoriety within the military or in the public media as "bad performers" in one way or another. Under these circumstances, it may have become easy, even fashionable, to be overly critical of these programs.

- Four of the programs were still wrestling with fundamental mission need or concept selection issues at milestone III. It is not clear whether these were legitimate unresolved questions which were compounded by inadequate T&E data or whether they were spurious issues which simply muddled the decisionmaking process.

- Three of the programs showed evidence of intra-service disputes over priority or disagreement regarding the adequacy and meaning of DT&E.

- Finally, four of the programs did not have an approved TEMP early enough in the FSD phase to significantly effect the adequacy of T&E data presented to the DSARC. In fact, all of the programs began FSD long before the TEMP was an established requirement. However, at least in one case, the absence of a TEMP was overcome by some other agreement which served a similar function.

EXHIBIT V-13

Common Factors in Programs Studied

	MCLWG	AEGIS	EF-111A	ROLAND	ALCM	PATRIOT
LONGER THAN AVERAGE	●	●				●
OUTSIDE ATTENTION	●		●	●		
MISSION NEED OR CONCEPT SELECTION ISSUE	●	●	●	●		
INTRASERVICE DISPUTES	●				●	●
LATE OR NO TEMP	●		●	●*		●

* OFFSET SOMEWHAT BY "POWERS REPORT" AND SUBSEQUENT PLANNING

More specifically, there were a number of underlying causes for inadequate T&E data which could be identified for each program, as shown in Exhibit V-14. These included:

- Changes in policy
- Shifts in emphasis
- Ambiguities in system scope
- R&D problems
- Test problems

Each of these is discussed below.

(1) Changes in Policy

A change in policy is a substantive change in the acquisition process as reflected by DOD directives or instructions which occurred since a program's last major milestone review. Included are such things as the objectives, activities, and guidelines for acquisition phases; the suporting documentation and procedures; and the criteria for moving from one phase to the next. As was pointed out in Chapter III, policy swings can be rather dramatic and acquisition strategies formulated at one point in time can be grossly unacceptable at another. This was the case in five out of the six programs studied. Of course, this problem is exacerbated by long development programs.

(2) Shifts in Emphasis

Shift in emphasis is similar in that it may have the same effect as a change in policy; however, it may or may not become official. It deals with the changing preoccupation of defense decisionmakers with a selected aspect of the acquisition process, such as design-to-cost, or operational suitability. Shifts in emphasis may result from the zeal of individual OSD staffers or from a heightened awareness of some particular issue within the acquisition community. Both the EF-111A and the Roland programs seem to have gotten caught by OSD emphasis that was different than when the programs were last reviewed.

(3) Ambiguities in System Scope

Inadequate T&E data frequently resulted from confusion or ambiguities regarding the scope of the system tested, i.e., how representative the test hardware was of the complete production configuration operating in a realistic environment with all its associated systems. There were numerous disagreements

EXHIBIT V-13
Causes of Inadequate T&E Data

	MCLWG	AEGIS	EF - 111A	ROLAND	ALCM	PATRIOT
CHANGES IN POLICY	●	●	●	●		●
SHIFTS IN EMPHASIS			●	●		
R & D PROBLEMS			●	●	●	●
TEST PROBLEMS	●		●		●	
AMBIGUITIES IN SYSTEM SCOPE	●	●	●		●	

at milestone III over the "reasonableness" of extrapolating what was actually tested to what was planned to be produced. This included:

- The ease or difficulty in moving from a partial to a full capability, most notably involving software and system integration.
- The realism of the test environment, including operation and maintenance by typically skilled personnel.
- The use of simulation as a substitute for missing components or systems.

Furthermore, confusion over the scope of the system to be tested may have reflected a situation in which the critical test issues were not well defined and, consequently, the extent and nature of the testing required was indeterminable.

It is clear that a fixed policy regarding the scope of systems to be tested could not have been applied across the board to all the programs studied because of differences in technical risks, priority, etc. However, questions regarding adequacy of T&E data due to system scope could have been reduced if the T&E program needed to address the critical test issues had been defined and agreed to in advance on a case-by-case basis. A TEMP would have provided a vehicle for such negotiation had it existed early enough to affect the test program.

(4) R&D Problems

R&D problems encountered during development, as part of the normal R&D process, sometimes consumed funds that were planned for T&E. The resulting cost constraints frequently led to an inadequate number of prototypes for training and testing and, sometimes, a prototype configuration that differed significantly from that intended for production (see system scope, above). R&D problems also caused slips which reduced the time available for T&E prior to a fixed milestone review. Milestone III in particular is sometimes viewed as fixed in time because of an institutionalized perception that appropriated funds cannot or should not be released until after DSARC approval. Since funding profiles are established years in advance as part of the PPBS process, and are often based on optimistic cost and schedule estimates, it is not uncommon for RDT&E money to be declining and production money to be available prior to completion of the

full-scale development phase. This puts an added pressure on the Service and OSD to hold the DSARC in spite of major deficiencies (as in the case of Patriot, for example) because of concern over losing the production funding.

The resulting schedule pressures led to inadequate test time, short cuts in the evaluation of reliability and maintainability, and in logistics supportability. It was not uncommon to have the T&E time planned in the original schedule compressed as the DSARC date approached. OT in particular was subject to compromise by R&D problems. Operational tests were reduced or omitted, OT was combined with DT or, in the worse case, hardware was released for OT before all developmental problems were resolved.

Fixing the milestone reviews as a result of lead time in the budget process is a mixed blessing; it sets aggressive goals (which may work to overcome Parkinson's law); however, it also encourages shortcuts in order to meet artificial deadlines. Additional funding flexibility during the R&D to production transition may be required to deal with this dilemma.

(5) Test Problems

Test problems which prevented acquiring adequate T&E data included the availability of prototypes for test, the availability of targets and other consumables, and problems with range facilities and instrumentation. Interestingly, range priority was not found to be a problem, probably because the major system studied all enjoyed a higher priority than the non-major systems with which they had to compete.

A related problem was that competitive prototyping, as in the ALCM program, diluted the T&E effort to the point where marginal data was available for review. This is a serious problem which must be evaluated in terms of the whole question of the value of competition in defense acquisition.

While cases of inadequate data an be attributed to the underlying causes described above, it also appears that the problem of inadequate T&E data is often dominated by milestone reviews by other decision drivers and that limited production is frequently the resulting compromise. There is circumstantial evidence that adequacy of

T&E sometimes is used as a convenient surrogate for short term affordability problems. At other times substantive question about risks and readiness for production are overlooked in defer- ence to urgency, momentum, national priority and other issues such as NATO RSI.

* * * * *

The decision profiles in this chapter showed that while specific circumstances varied, the programs had similar attributes, there were common underlying causes of inadequate T&E data, and that frequently the milestone decision was dominated by other, non-T&E related, factors. The next chapter integrates these and earlier findings into a set of conclusions and recommendations which address the study objectives.

VI. CONCLUSIONS AND RECOMMENDATIONS

VI. CONCLUSIONS AND RECOMMENDATIONS

During the period under study, the majority of milestone II and III decisions were made conditionally, and the frequency of conditional decisions has increased steadily since 1977. Use of "limited production" at milestone III is now commonplace and indicates a movement toward a five-milestone system as a means of incremental confidence building.

Conditional decisions can add to the time and cost required to develop new weapon systems; however, they can also prevent large-scale mistakes, for example, by restricting capital investments until risks have been minimized. There is no such thing as the "correct" balance between caution and expediency--it changes as the needs and the resources on the defense establishment change. Similarly, it is futile to search for an absolute definition of adequate T&E data. It is more constructive to attempt to understand the way in which T&E data are integrated into the defense decisionmaking process and to make recommendations for improving that process.

1. CONCLUSIONS

The following conclusions are drawn in response to the questions that were originally posed in Chapter 1.

(1) Inadequate T&E Data Contribute to Some Conditional Decisions

Inadequate T&E data as perceived by the DSARC principals are one reason for conditional decisions. However, instances of inadequate T&E data occur relatively infrequently and usually in combination with other reasons, such as failure to meet a performance threshold. In any case, milestone III decisions are commonly dominated by other factors such as the urgency and national priority of the program, the momentum of the program, the long lead times associated with production, and the desire to preserve budgeted or appropriated funds. Furthermore, the adequacy of T&E data appears sometimes to be used as a convenient substitute for near-term affordability problems.

Thus, inadequate T&E data cannot be singled out as the primary cause for the conditional decisions studied. Nevertheless, it would be desirable to eliminate individual instances where the DSARC imposes conditions which extend the acquisition cycle because they question the adequacy of the T&E data presented to them.

(2) Underlying Causes of Inadequate T&E Data Are Both Internal and External to the Program

Adequacy of T&E data is a difficult concept to define precisely. However, it is clear that it is a function of the expectations of the DSARC principals and that these expectations are ill defined and constantly changing in response to program-unique factors and external events. Further, the principals themselves change from review to review of a given program because the R&D cycle is, in general, longer than the political cycle. This makes establishing a stable framework for conducting individual program reviews difficult and opens the larger issue of how to institutionalize centralized decisionmaking in a decentralized operating environment.

When the DSARC concludes that there are inadequate T&E data on which to base an unconditional decision, it is generally as a result of a lack of communication between the Service and OSD. Either the expectations of the principals were never established or there was a change in the program or the decision environment which was not communicated to the other party. In the six cases studied, the underlying causes of inadequate T&E data most frequently included:

- Changes in acquisition policy since the last milestone review
- Shifts in emphasis by key players or groups
- Ambiguities regarding the scope of the system to be tested, including identification of the critical test issues
- R&D problems which reduced the time and/or funds available for T&E prior to a "fixed" milestone review which is perceived as necessary to preserve follow-on funding
- Problems in conducting the tests themselves due to difficulty in obtaining test resources or to range/instrumentation limitations.

Only the last two are even partially within the control of the Program Manager or developing Service. The others primarily relate to the specific expectations of the DSARC principals regarding the quantity and quality T&E data to be presented at milestone reviews.

(3) There Is Not Always Effective Communication Between OSD and the Services With Respect to the Quantity and Quality of T&E Data That Is Required for Milestone Decisions

Formal communication between OSD and the Services regarding T&E did not routinely exist until about 1979 when the TEMP began to be applied rigorously as a means of establishing a "contract" as to the quantity and quality of T&E required to gain approval to proceed to the next acquisition phase. In a sense, according to present policy, an approved TEMP declares and affirms on a case-by-case basis the expectations of the DSARC principals in an attempt to bring stability to the decisionmaking process. However, it is not yet clear that future decisionmakers will (or should) be bound by such agreements as program and policy users change and new facts are brought to light. Thus, as important as the concept of the TEMP as a contract, is its use to provide continuing communication between the parties regarding test plans that realistically reflect the acquisition strategy--including the scope of the system to be tested, the DSARC acceptance criteria, and the existing funding profile. Since 1 to 2 years will be required for programs to respond to the policies begun in 1979, it is not yet known whether they will be effective in avoiding the T&E inadequacies that were observed.

(4) The Way in Which Adequacy of T&E Is Addressed in the Present Milestone Review Process Needs Improvement

The independent T&E assessment of the DDT&E is an effective means of evaluating the results as well as limitations of tests conducted prior to a milestone review. However, it is but one part of the overall evaluation that must occur in reaching a milestone decision. A judgment regarding the adequacy of T&E data cannot be made out of context of a total risk assessment which considers the costs of delay, the military urgency, and an assessment of the consequences of success or failure of the alternative courses of action. Such a total risk assessment is beyond the charter of the DDT&E and it is not clear that it should be delegated to him. Presently, overall evaluation, including risk assessment, is done by all the DSARC principals by means of the lengthy DCP coordination and prebrief process. While this process has the attribute of consensus building within OSD, it is not clear that it is the most efficient way to generate an independent overall evaluation for consideration by the DSARC.

It is also observed that the DSARC spends a disproportionate amount of time and energy on the milestone III review at which point they have relatively little opportunity for altering either the system concept or the acquisition strategy in the face of enormous pressures to proceed with production. Furthermore, many of the questions regarding the adequacy of T&E data arise out of the immaturity of the hardware subjected to the DT&E and IOT&E conducted prior to the major production decision. Present policy states that milestone II includes authorization for limited production of hardware for IOT&E. However, there appear to be a number of misconceptions and barriers which have prevented full implementation of this approach. If this policy were strictly adhered to, it would reduce some of the questions regarding adequacy of T&E prior to a rate production decision. It would also further increase the importance of the milestone II review.

(5) Institutional Pressures Pose Obstacles to The Developer and Tester in Planning and Conducting an Adequate T&E Program

Institutional pressures, including competion for resources, cause program planning to be success oriented. This tends to minimize the flexibility of program managers to deal with R&D problems as they occur and it frequently forces compromises in the adequacy of planned test programs. Furthermore, resource perturbations which result from the PPBS process are not translated into changes in acquisition strategies. PMs, trying to do more with less, invariably run out of time or money trying to generate T&E data prior to a fixed milestone review.

Current practice under the Uniform Funding Policy, which allocates costs at major ranges and test facilities to individual projects rather than from an institutional fund, aggrevates this problem and may inhibit comprehensive, thorough testing under some circumstances.

2. RECOMMENDATIONS

Exhibit VI-1 shows the recommendations of this study displayed against the conclusions just discussed. Each recommendation is described below.

(1) The DSARC Should Place Greater Emphasis on Earlier Milestones in the Acquisition Process, Particularly on Milestone II

It is clear from the cases studied that the DSARC principals were placing a disproportionate emphasis on

EXHIBIT VI-1

Conclusions and Recommendations Matrix

CONCLUSIONS

RECOMMENDATIONS	#1 INADEQUATE T&E CONTRIBUTES TO SOME CONDITIONAL DECISIONS					#2 UNDERLYING CAUSES ARE BOTH INTERNAL AND EXTERNAL TO THE PROGRAM			#3 COMMUNICATION BETWEEN OSD AND THE SERVICES NEEDS IMPROVEMENT		#4 THE MILESTONE REVIEW PROCESS NEEDS IMPROVEMENT			#5 INSTITUTIONAL PRESSURES POSE OBSTACLES TO THE DEVELOPER AND TESTER		
	OCCURS LESS FREQUENTLY THAN PERFORMANCE BREACHES ETC	SOMETIMES USED AS A SURROGATE FOR AFFORDABILITY ISSUES	OTHER FACTORS TEND TO DOMINATE AT MILESTONE III	DISPARATE TO ELIMINATE FUTURE INSTANCES	EXPECTATIONS ARE NOT DEFINED AND CONSTANTLY CHANGING • CHANGES IN POLICY • SHIFTS IN EMPHASIS • AMBIGUITIES IN SCOPE	PROGRAM RELATED CAUSES INCLUDE • OPTIMISTIC SCHEDULES • R&D PROBLEMS T&E SHORTCUTS • TEST RESOURCES AND RANGE LIMITATIONS	CRITICALLY IMPORTANT IN A DECENTRALIZED FRAMEWORK		NEW TEMP PROCEDURES HAVE NOT HAD TIME TO WORK	BETTER OVERALL EVALUATION IS NEEDED	DISPROPORTIONATE EMPHASIS ON MILESTONE III	LIMITED PRODUCTION NOT USED EFFECTIVELY	SUCCESS ORIENTED PLANNING	FIXED MILESTONES DUE TO LORD IMPLICATIONS	RESOURCE PERTURBATIONS NOT COUPLED TO PLANS	ALLOCATION OF RANGE COSTS
(1) PLACE GREATER EMPHASIS ON EARLIER MILESTONE REVIEWS	●		●							●						
(2) CLARIFY THE POLICY REGARDING MEANING OF MILESTONES II AND III	●			●						●	●					
(3) DEVISE NEW RESOURCE PROGRAMMING TECHNIQUES TO FOSTER TRANSITION FROM R&D TO PRODUCTION				●		●							●			
(4) OBTAIN BETTER CONSENSUS REGARDING OPERATIONAL ENVIRONMENT ON A CASE BY CASE BASIS				●	●		●					●				
(5) APPLY MORE RIGOROUS T&E PLANNING REGARDING SYSTEM SCOPE				●	●	●	●					●				
(6) EMPHASIZE USE OF THE TEMP AS A MEANS OF STRUCTURED COMMUNICATION	●			●	●	●	●	●				●		●		
(7) INSTITUTE MORE RIGOROUS ANALYSIS OF ALTERNATIVE COURSES OF ACTION IN THE CONTEXT OF AN OVERALL EVALUATION	●	●	●	●					●							
(8) REQUIRE IMPACT STATEMENTS FROM PM'S FOLLOWING REVISION TO MAJOR ACQUISITION DIRECTIVES, AND BE MORE SENSITIVE TO DOCUMENTING CHANGES IN EMPHASIS AS IT OCCURS				●	●		●									
(9) CHARGE DDT&E WITH RESPONSIBILITY FOR INSTITUTIONAL IMPROVEMENTS IN T&E AFTER EACH DSARC				●												●
(10) CONTINUE TO MONITOR THE RELATIONSHIP BETWEEN T&E DATA AND DSARC DECISIONS				●			●	●	●	●	●					
(11) DETERMINE THE UNDERLYING CAUSES OF PERFORMANCE BREACHES	●											●		●		

the Milestone III review--at which they had no realistic alternative other than to place the system in some form of production irrespective of the technical or economic risks. The emphasis on Milestone III is due partly to the magnitude of resources that are apparently being committed at that point; to the interest and involvement by the Congress in the production decision; and to the tendency of the decisionmakers to focus on rather near term decisions.

From a practical standpoint, a better balance between available information and the ability of the bureaucracy to influence program direction occurs earlier in the program about the time of Milestone II. This is where the major commitment to the program is, in fact, made and this is where significant adjustments to the acquisition strategy can be made without major economic or schedule disruptions. Furthermore, this is about where the critical transition in terms of the magnitude of resource expenditures occurs. While a decision at milestone II necessarily must be made without extensive developmental testing (and hardly any meaningful operational testing), it is not too early to methodically identify the critical test issues or to conduct a careful analysis of the test and evaluation plans.

Shifting the DOD emphasis to Milestone II will not be easy; however, policy pronouncements followed by increased vigor and intensity of Milestone II reviews will begin the necessary acculturation.

(2) <u>The Policy Regarding the Meaning of Milestones II and III Should be Clarified</u>

Requiring T&E data on hardware that is representative of the production configuration prior to the production decision is a classic "Catch 22." Present policies address this problem by authorizing initial production for the purpose of initial operational test and evaluation at milestone II and full scale (or rate) production at milestone III. However, current practice appears to segregate the R&D and production processes to the point that multiple milestone III reviews are required, i.e., a five milestone system is used which includes a separate limited production as well as a rate production decision.

If two separate production reviews are desired, then the appropriate directives should be changed to reflect current practice. However, if authority for initial production for IOT&E is to be granted at milestone II, then the present policy should be restated and emphasized. Through policy issuances, the Services should be encouraged to develop acquisition

strategies that utilize early production hardware for IOT&E and that require authority for rate production only at milestone III.

(3) New Resource Programming Techniques Should be Devised to Deal With the Critical Transition From RDT&E to Production Funding

In the cases studied, the Services (and by inference the DSARC) frequently felt compelled to hold a Milestone III review rather than risk losing production funding that was programmed 3 to 5 years earlier. The fixed milestone date, coupled with generally optimistic planning estimates and normal R&D problems, tended to cause shortcuts in the DT&E and IOT&E needed to support fully a recommendation for rate production. Since no means has been found to eliminate either optimism or R&D problems, added flexibility is required to adapt the acquisition process to the realities of the PPBS process.

This problem has both cultural and regulatory dimensions. Differing organizations and procedures within the DOD, and to some extent within the Congress, are used to manage and control the RDT&E and production appropriations. The chasm that sometimes exists between the two groups and the general lack of interchangeability of the two types of funds tends to cause program managers and Service proponents to make inappropriate compromises in T&E programs and to participate in budgeting "gamesmanship" rather than deal with the realities of a given situation.

If Milestone III is to be seriously viewed as the decision for rate production, then some relief must be found from the triangle consisting of optimistic schedules, inevitable R&D problems and milestone review dates which are driven by a desire to preserve follow-on funding. New management techniques regarding R&D and production funding would allow graceful slips in milestone III reviews without loss of funding until a legitimate case could be made for rate production.

(4) Greater Care and Attention Should be Given to Defining the Meaning of the Term "Operational Environment" within the Context of Each Major System Acquisition.

Throughout this study of DSARC decisionmaking, a persistent issue arose as to the expectation of the decision participants over the requirement to test

systems in their "Operational Environment." In some cases technical issues were involved such as numbers of Air Defense Systems or the complexity of the electromagnetic environment in which a system was tested. In other instances it involved the use of contractor personnel to operate, maintain and repair the system under test. Currently employed terminology such as "as near as possible to the operational environment" is hardly susceptible to universal interpretation. Reaching an early consensus on the definition of the operational environment on a case-by-case basis would permit better T&E planning and eliminate future misunderstanding. (See also recommendation #5 regarding disciplined control of the scope of the system to be tested.)

(5) More Rigorous System Engineering Discipline Should be Used to Establish the Scope of the System to be Tested and to Control the Impact of Changes to Hardware and Software Configurations and Test Programs

A persistent issue between the services and the OSD decisionmakers is the scope of the system that was tested, i.e., how much of the entire system, including all its subsystems and related or associated systems, was or should be tested? In what degree of final configuration and in how realistic an environment? Downstream controversies, particularly at milestone III, could be minimized by means of system block diagrams, work breakdown structures or other system engineering tools to depict the relationship of planned tests on the system in its entirety. By arraying the planned test configuration against the total system in a systematic way, tradeoffs between cost, schedule and risk could be realistically assessed and future ambiguities avoided. On larger and more complex programs the TEMP could be brought under formal configuration management control. In the event that downstream resource allocations required readjustments of the program, inadvertent departures from established expectations could be identified for coordination with DDT&E.

(6) The TEMP Should Be Viewed as the Primary Mechanism For Structured Communication Between OSD and the Services Regarding the T&E Data Required to Progress From One Phase of Acquisition to Another

The TEMP can be viewed as a contract between OSD and the Services in which the adequacy of T&E data at each phase of the process is defined. However, it is important to recognize that the contract legitimately

may be modified by either party: the program manager has an obligation (consistent with his reponsibility for efficient execution of the program) to respond to changing situations as necessary, and OSD must focus on making good decisions in the light of existing circumstances regardless of previous commitments. Thus, the real value of the TEMP is as a means of imposing stuctured communications between the parties. This means that the TEMP should be updated whenever changes occur in any of the fundamental underlying assumptions on which the T&E program is based, such as policy, funding profile, threat, etc.

(7) Greater Emphasis Should be Placed on Identifying and Analyzing Alternative Courses of Action in the Context of an Overall Evaluation

A singular weakness found in the cases studied was the lack of a clear cut responsibility for bringing together the many factors and issues which require careful balancing for informed decisionmaking. The present process depends on the coordination of the DCP to reach a consensus--a process that is repetitive, time consuming and sometimes results in a recommendation that represents the least common denominator rather than the optimum course of action.

There are alternatives to the DCP as a means of conducting an overall evaluation. Responsibility for this function could be assigned to an existing or new member of the Defense Acquisition Executive's staff, or it could be delegated directly to the program manager who, in general, has greater expertise in the specific program under review. In any case, an overall evaluation would include technical, affordability and other issues of which the independent assessment of the DDT&E is but one part. Consequently, the DDT&E assessment, even if it is a preliminary or partial one, should be made early enough to be seriously studied in an integrated review of all factors.

Finally, an overall evaluation should, as a matter of course, identify the costs and other ramifications of each of the alternative courses of action, particularly those that impose delays or result in uneconomic production. The costs of such decisions must be identified and consciously acknowledged as part of the decisionmaking process.

(8) Impact Statements Should be Required From Program Managers of Major Systems When Changes are Made to Defense Acquisition Policy

One finding of this study was the extent to which the developmental time of major systems exceeded the duration of the acquisition framework under which they were carried out. Often, policy changes were not fully reflected in programs when they came up for milestone review, partially because resources were not provided to comply with the policy change.

Coordination of "for comment drafts" is one way to surface such problems but this approach apparently has not been effective in avoiding downstream controversy--either because the questions do not get to the PMs themselves, or because their answers are buffered by the Service headquarters staffs. An institionalized process whereby PMs prepare specific statements on the impact to their programs of changes to high level acquisition directives would deal more effectively with problems of retroactive application of policy and requirements for additional resources than when deferred to the next POM or ignored altogetner. Such statements could be followed by or made part of Program Change Requests which would bring the effected programs into compliance.

A similar but more difficult problem is the impact of changes in emphasis by defense decisionmakers which has the effect of policy but which is never formally communicated to the Services. In these cases, OSD can only be more sensitive to documenting defacto changes in policy as they occur.

(9) The DDT&E Should Make Recommendations Regarding Institutional Improvements in T&E as Part of the DSARC Process

Although the DT&E has important responsibilities for planning, coordinating and budgeting for national ranges and test capability, the "lesson learned" from the T&E experiences of a given program while fresh within the context of a particular DSARC review are not immediately fed back into the institional management process. The independent technical evaluations provided for the DSARC are important, but they do not help in managing the department as an institution. Contemporaneous recommendations for improving the T&E process would deal more effectively with the broader institutional issues such as the impact of T&E funding policies on system acquisition and the need for new or improved test capabilities.

(10) Continue to Monitor the Relationship Between T&E Data and DSARC Decisions

The December 1979 revision to DODD 5000.3 and the current practices of the DDT&E encourage a more formal, structured dialog between OSD and the Services regarding the adequacy of test planning and data as recommended in this study. However, since it takes several years for a program to evolve in response to policy initiatives, it will not be known how these policies effect the frequency of conditional decisions until about 1982 or 1983. Consequently, the issue of adequacy of T&E data and conditional decisions should be monitored to determine whether structured communication via the TEMP is effective.

(11) Determine the Underlying Causes of Breaches of Performance Thresholds

Most of the conditional decisions that occured during the period under study could be attributed, at least in part, to a failure to meet one or more performance thresholds. An investigation of the underlying causes of this phenomena was beyond the scope of this study; however, there is a large potential payoff in gaining an understanding of how and why performance breaches contribute to increases in the time and cost of system acquisitions.

Failing to meet a perfomance threshold in nearly 90 percent of the DSARCs that resulted in conditional decisions indicates more systemic problems than can be explained simply by R&D uncertainties. Either performance specifications are being set unrealistically high, i.e. there is an institutional desire to achieve only quantum advances in technology, or milestone reviews are being held too early in the acquisition cycle to allow for the desired degree of system maturity, or the DSARC is failing to appreciate the developmental nature of the systems under review. Techniques such as performance growth curves and preplanned product improvement address these problems, however a detailed study of the underlying causes of performance breaches is required.

APPENDIX A

BIBLIOGRAPHY

BIBLIOGRAPHY WITH SELECTED ANNOTATIONS*

I. OMB CIRCULARS, DOD DIRECTIVES, INSTRUCTIONS, AND SERVICE REGULATIONS

A. ACQUISITION

- Circular A-109, "Major System Acquisition," April 5, 1976.
- DOD Directive 5000.1, "Major System Acquisitions," March 19, 1980.
- DOD Directive 5000.2, "Major System Acquisition Procedures," March 19, 1980.
- DOD Directive 5000.39, "Acquisition & Management of ILS for Systems and Equipment," January 17, 1980.

B. TEST AND EVALUATION

Department of Defense

- DOD Directive 5000.3, "Test and Evaluation," December 26, 1979.
- DOD Directive 3200.11, "Use, Management and Operation of DOD Major Ranges and Test Facilities," September 29, 1980.
- DOD Instruction (Draft), "Operational Suitability Requirements and Assessment," August 22, 1980.

Army

- AR 10-4, "US Army Operational Test & Evaluation Agency."
- AR 70-10, "Test & Evaluation During Development and Acquisition of Material."
- AR 71-3, "User Testing."

---

* In most cases synopses have been taken in whole or in part from original work. Entries are shown chronologically within each section.

- AR 700-78, "Production and Post-Production Testing of Army Material."

- AR 70-64, "Joint Service Interface on DT&E," May 15, 1979.

Navy

- OPNAVINST 4720.9, "Approval of System and Equipment for Service Use."

- OPNAVINST 3960.10, "Test and Evaluation," July 31, 1980.

- NAVMATIST 3960.6, "Test and Evaluation."

- NAVMATINST 3960.7, "Test and Evaluation of Ship Acquisiton."

- NAVMATINST 3960.8, "Land-Based Test Site (LBTS) Selection Policy."

- NAVSEA 0900-LP-095-2010, "Ship Construction Test Fitting Out, and Shakedown of U.S. Naval Vessels and Trials Manual."

- NAVMATINST 3970.1, "Joint Service Inderface on DT&E," May 15, 1979.

Air Force

- AFR 80-14, "Test and Evaluation" and AFSC Supplement 1, February 19, 1981.

- AFSCR 172-8, "Test and Evaluation Support."

- AFSCR 23-26, "Air Force Test and Evaluation Center (AFTEC)."

- AFSCR 80-24, "Joint Service Interface on DT&E," May 15, 1979.

Marine Corps

- MCO 5000.11A, "Testing and Evaluation of Systems and Equipment for the Marine Corps," July 2, 1979.

- MCO 5000.2, "Joint Service Interface on DT&E," May 15, 1979.

C. MISCELLANEOUS

- DOD Directive 4155.1, "Quality Programs," August 10, 1978.

- DOD Directive 5000.40, "Reliability and Maintainability," July 8, 1980.

- DOD Instruction 7000.2, "Performance Measurement for Selected Acquisitions," June 10, 1977.

II. STUDY REPORTS AND ARTICLES

A. TEST AND EVALUATION (GENERAL)

- Compendium of Test Terminology, December 1978. Compiled by the Joint Logistics Commanders Ad Hoc Group on Test and Evaluation Planning Guidance.

- Hobbit, Jerry N., Air War College Research Report No. 372, "T&E the Persistent Problem," April 1978.

- Defense Systems Management Review, "Test & Evaluation," Vol. 1, No. 5, Winter 1977, complete issue including the following articles:

  - Lt. Gen. Lotz, Walter E., "T&E in the Department of Defense"

  - Maj. Gen. Powers, Patrick W., "Streamlining Army Testing"

  - Maj. Gen. Leaf, Howard W., "Improving AF Independent Operational Testing"

  - Com. Donovan and Lt. Com. Fitzgibbons, "OT&E of Ships: Policy and Practice"

  - Cpt. Sherman T.H., "T&E in Defense Procurement"

  - Gleghorn George G., "Optimizing Test Vehicle Test Programs"

  - McCord John W., "Management of Major DOD Test Facilities"

  - Lt. Col. Washington D.L., "T&E of TRI-TAC Equipment"

  - Kreiner H.W., "How Much Testing is Enough?"

  - Karns, Charles W., "Test and Evaluation Policy"

  - Lt. Col. Butchho M. T., "Aircraft OT&E"

  - Lt. Col. Fritz, Martell D., "Use of Training Exercises for T&E"

- Love, D.T., "Challenges in Aircraft Engine T&E Programs"

- Col. Swett, B.H., "Reliability & Maintenance for the Acquisition Process"

- Rusterburg, Owe-and Geese, "B-1 Structural Test Program"

- Lt. Col. Mache R.B., "Product Liability."

. DSMC, Schug R.J., "DT&E vs OT&E for Naval Air System Command Weapon System Acquisition," Study Project Report (DTIC A043212, May 1977).

SYNOPSIS: The purposes of this study are: To review current directives, reports, factors, and criticisms that determine the current structure and methodology for test and evaluation within the Navy; to determine the existence of overtesting or inadequate testing, and the emphasis on Operational Test and Evaluation, which affect the acquisition of weapon systems in the Naval Air Systems Command; and to examine the extent and effectiveness of integrated testing.

. ODDR&E/Defense Science Board, "Report of Task Force on T&E Policies," February 17, 1977.

SYNOPSIS: An earlier Defense Science Board Task Force on Test and Evaluation, established in 1972, had summarized and delineated general guidelines for Department of Defense review and monitoring of the test and evaluation aspects of development programs. This report primarily considered such questions as: Is testing taking too much time? Are we overtesting? Is all of the testing worthwhile? Are we establishing another bureaucracy?

The Task Force concentrated on T&E policies by examining a small group of representative programs which have paid attention to test and evaluation from the office of the DD/T&E. They concluded that T&E is a fundamental and integral part of current system acquisition procedures. Further, they found little or no overtesting being done; what testing was done contributed to the improvement and verification of system performance.

. DSMC, Horik, T.H., "The Navy T&E Process in Major Systems Acquisition," Study Project Report (DTIC A035935), November 1976.

SYNOPSIS: Test and Evaluation (T/E) is an integral and continuing part of the weapons system acquisition cycle. In the process of gaining knowledge from research through development and production, some T/E occurs each step of the way. In the past, it has often been compromised as attempts were made to meet fixed deployment dates. This study investigates the Navy's T/E process in the procurement of ships and aircraft, including some recent changes in T/E concepts brought about by recommendations of the Blue Ribbon Defense Panel and DOD Directive 5000.1. The unique problems associated with low-risk, long-duration, conventional hull ship acquisition programs are discussed including Land Based Testing, Initial Operational Test and Evaluation (IOT/E), and DCP/DSARC alternatives. Similar problems associated with high-risk but relatively short aircraft acquisition programs are also addressed with emphasis on concurrency, prototyping and OT/E. Incorporation of partial solutions to these problems (as evidenced in the Patrol Frigate and F-14 programs) are examined for applicability to future T/E programs.

ODDR&E/Defense Science Board, "Report of Task Force on T&E," April 2, 1974.

SYNOPSIS: The Defense Science Board Task Force on Test and Evaluation was established at the request of the Director of Defense Research and Engineering, on behalf of the Deputy Director (Test & Evaluation) to develop guidance on test and evaluation through examination of a group of representative weapon systems acquisition programs.

The resulting report presents guidance on T&E at two distinct levels. At the most general level, the report discussed a number of issues which are appropriate for all weapon systems acquisition programs, and are generally matters of basic policy. These issues are: reliability, computer software, human factors, the "T&E Gap", functional specifications versus design specifications, offense/defense testing, portable instrumentation, ship testing, and test planning. In addition, the report includes a general checklist of items, generally applicable to all systems development and deployment, which is organized for a rapid overall review of T&E aspects.

- DSMC, Bryder W.D., "The T&E Evolution (Relative to USAF)," Study Project Report, DTIC A040374, May 1974.

  SYNOPSIS: The Blue Ribbon Defense Panel Report of July 1970 provided the impetus for major changes in the structure and conduct of test and evaluation (T&E) in the Department of Defense. This report traces the actions taken by OSD, Air Force and Systems Command over the last four years up to the formation and implementation of the Air Force Test and Evaluation Center (AFTEC) in January of 1974. To ascertain the probable direction of T&E, several members of the T&E community were interviewed. The results of these interviews are presented and discussed. They indicate a uniform and stable policy throughout the Air Force and an increasing role of AFTEC in the T&E business as they become active and involved in the emerging programs. No major changes are anticipated in the current T&E policies and directives.

- GAO Report, "The Importance of T&E in the Acquisition Process for Major Weapon Systems," August 1972.

  SYNOPSIS: This report examined T&E on 13 weapon systems. It concluded that most systems did not have adequate test plans, T&E was not accomplished in a timely manner, T&E data was not available prior to those times in the acquistion cycle when decisions had to be made. The report also addressed waiver problems. The GAO recommended completion of T&E prior to key decisions, adequate controls over T&E waivers and improved reporting of T&E results. Some believe the report contributed to the "T&E explosion."

B. <u>DEVELOPMENTAL T&E</u>

- GAO Report PSAD-79-86, "Effectiveness of DOD DT&E," June 25, 1979.

  SYNOPSIS: In 1978, GAO reviewed the military services' development test and evaluation (DT&E) of six weapons systems totaling an estimated $12 billion in development and procurement costs. Inadequate testing is principally due to the expense, the difficulty of staging and execution, the time consumed, and belief that negative test results may jeopardize or delay a program. To overcome the unenthusiastic attitude toward testing, the type of results expected from the

testing must be defined. Possibly, a strong testing activity is the only means of accomplishing this objective. The Department of Defense spelled out testing procedures in a 1971 policy.

This study found that, as design and development phases progressed, the programs experienced cost growth, schedule slippages, and a reduction in performance levels. To remedy these effects, the responsible agencies, without formal assessment of risks and benefits, reduced the scope of, and hardware for, DT&E during critical tests; advanced programs into operational testing and production before completion of DT&E; and approved concurrent development and production; all without formal assessment of risks and benefits. In five of the six programs, planned tests for resolving technical uncertainties either were reduced or not performed, and final test reports either were unavailable or omitted the critical issues to be resolved before a given development phase could be considered complete. These practices could result in the systems entering production before their performance capabilities are known. Judging from past experience, this is an unsatisfactory procedure because systems have emerged from testing with degraded performance or with additional costs incurred in raising performance levels to acceptable standards. Actions which would reduce development time and costs include the elimination or reduction of development tests of systems and subsystems, the combination of DT&E's with initial operational test and evaluation (IOT&E), and a clear assessment of the risks involved in concurrent development and production.

The recommendations of this report to the Agencies are that the Secretary of Defense should require that DT&E reports identify any deleted, reduced, or substituted tests, and state clearly the risks and implications of these actions. The Secretary should also require test plans to show the additional risks of joint DT&E/IOT&E test schedules and allow time to correct discovered deficiencies before proceeding to the next testing phase. Formal assessment of risks and benefits should be required before approval of concurrent development and production, and developers should be ordered to monitor sensitive system elements and high-risk areas closely, and provide schedule and funding reserves to resolve problems as they are identified.

C. OPERATIONAL T&E

- GAO Report PSAD-80-2, "Army Operational Test & Evaluation Needs Improvements" (CONFIDENTIAL), November 13, 1979.

- GAO Report PSAD-79-1, "Follow-on OT&E of Weapon Systems," October 19, 1978.

  SYNOPSIS: A survey of practices and procedures for follow-on operational testing and evaluation of weapon systems by the military services indicated that follow-on tests have been conducted adequately. However, cases were noted where critical testing was deferred until after a production decision was made, and full-scale operational testing has been conducted after weapon systems have been produced and deployed. In the Army, critical tests needed to support a low-rate production decision were deferred until after production units were available. In the Air Force most operational testing prior to the initial production decision is combined with testing performed by developers; this does not constitute sufficient operational testing. Unlike the Army and the Air Force, the Navy usually performs sufficient operational testing before production decisions, but the Navy also continues to conduct full-scale operational testing after systems have been deployed. Some operational testing of complex weapon systems will be necessary after production decisions are made; this could occur when there are changes in threat, major system modifications, or changes in doctrine or tactics. The Secretary of Defense should: assure that operational tests critical to determining system effectiveness and suitability are accomplished prior to initial production decisions; and, in cases where systems have undergone full operational test and evaluation, restrict additional testing to cases of significant changes in mission, threat, tactics, or system modifications.

- GAO Report PSAD 78-131, "Operational Test and Evaluation of Foreign Built Weapon Systems," July 25, 1978.

- GAO Report PSAD-78-77, "Navy OT&E: A Valuable Tool Not Fully Utilized" (CONFIDENTIAL), March 29, 1978.

SYNOPSIS: The Navy's Operational Test and Evaluation Force tests and evaluates Navy weapon systems in an operational environment and develops procedures and tactics for their use. The Test Force is separate and independent of developing, procuring, and using commands, and reports results of operational tests and evaluations directly to the Chief of Naval Operations.

The Navy has made considerable progress in planning and conducting tests and reporting test results. However, in some instances, Navy officials made procurement decisions before sufficient operational test and evaluation was completed, and some systems entered production before it was known whether they could perform their missions under combat conditions. Operational test and evaluation results completed on some weapon systems after production began showed significant performance problems. Deviations from prescribed testing procedures should be made only when the need for the system is critical or the risk of performance problems is minimal. In some cases, a complete operational test and evaluation could not be made because of limited test resources.

Recommendations of this report to Congress are that the Secretary of Defense should demonstrate through operational testing that weapon systems can perform their missions in typical combat environments before they are approved for production and see that the Navy's Operational Test and Evaluation Force is given adequate resources to perform the necessary test and evaluation.

D. T&E GUIDELINES

. As an outgrowth of the Defense Science Board Task Force on Test and Evaluation, the DD/T&E in April 1974 compiled a series of "T&E Guidelines" for various categories of weapon systems. These included:

- T&E Guidelines for Missile Weapon Systems
- T&E Guidelines for Aircraft Weapon Systems
- T&E Guidelines for Ship Systems
- T&E Guidelines for Ground Vehicle Systems

- T&E Guidelines for ASW Systems
- T&E Guidelines for Airborne ECM Systems
- T&E Guidelines for Airborne General Surveillance Radar Systes
- T&E Guidelines for Command and Control Systems
- T&E Guidelines for Common Test Gear.

E. OPERATIONAL SUITABILITY

. OUSDRE - DDT&E/OASDMRA&L/WS, "Operational Suitability Verification Study, Phase I Final Report," January 1980.

SYNOPSIS: This report presents the results of a study of the weapon system T&E process relating to operational suitability. The study developed policies and procedures which aid in early formulation of operational suitability requirements. Included were planning of comprehensive and effective test programs, which exploit the time and resources made available for testing, and accurate assessment of system operational suitability characteristics.

. GAO Report PSAD-81-17, "Effectiveness of US Forces can be Increased Through Improved Weapon System Design," January 29, 1980.

SYNOPSIS: This report discusses weaknesses in the Department of Defense's major system acquisition process which are contributing to the operation, maintenance, and support problems of many deployed weapon systems. This review was undertaken in an effort to identify changes which are needed in the acquisition process to minimize problems with future weapon systems.

F. MAJOR SYSTEM ACQUISITION

. GAO Report PSAD-79-106, "Review of the DOD's Implementation of Procurement Reforms," September 25, 1979.

. GAO Report PSAD-79-80, "Recommendations of the Commission on Government Procurement: A Final Assessment," May 31, 1979.

SYNOPSIS: As Federal procurement expanded, and individual problems multiplied, a patchwork of laws, directives, and regulations was installed to cope with the problems. The Congress, after exhaustive hearings in the late 1900's, created the Commission on Government Procurement to devise fundamental improvements. This report is the eighth status report on the recommendations proposed by the Commission on Government Procurement in December 1972.

- GAO Report PSAD-79-9, "Observations on OMB Circular A-109, Major System Acquisition by DOD," February 20, 1979.

- OUSDR&E/Defense Science Board, "Report of the Acquisition Cycle Task Force," March 15, 1978.

SYNOPSIS: This Defense Science Board Summer Study Task Force on the Acquisition Cycle examined a number of major systems acquisition case histories, changes in acquisition policy over the past two decades, and the actual workings of the program advocacy and budgetary processes. The Task Force concluded that, over the past 15-20 years, the acquisition process--from initial program concept to initial operational capability--has increased in length by approximately five years on the average. Most of this increase has been in the "front end" of the process, from the time of initial concept studies to the start of Full Scale Development (i.e., DSARC II) and at the "back end" from the completion of FSD at DSARC III to the achievement of a full operational capability (i.e., the production and deployment phase).

The Task Force also concluded that full-scale prototyping often introduces costly and unnecessary program delays; that the policy of "no concurrency" is being applied too rigidly and is inefficient and costly in many cases; and that sequential and separate testing systems by developers and by the users (instead of the independent evaluation of jointly obtained test data) often adds unnecessary expense and time. It was also concluded that the practice of incremental decisionmaking by requiring the achievement of a consensus between the advocates and the antagonists at each DSARC milestone is a cumbersome process that tends to prolong programs so much that--during the acquisition cycle--the threat, the players, and even the concepts of

warfare often change, resulting in the necessity to cancel programs and start new ones to meet the threat, with the concomitant waste of limited national resources.

Finally, the Task Force expressed concern that without very careful attention to the implementation of OMB Circular A-109, this well-intentioned policy could very likely give rise to institutionalized procedures that will significantly increase the length of the "front end" phase of program acquisition rather that shortening it as intended and without achieving the intended degree of program flexibility and stability in acquisition policy.

- GAO Report PSAD-78-14, "A Critique of the Performance of the Defense Systems Acquisition Review Council," January 30, 1978.

G. CONGRESS

- GAO Report PSAD-80-37, "SARs: Defense Department Reports That Should Provide More Information to Congress," May 9, 1980.

  SYNOPSIS: This report points out ways in which the Department of Defense's selected acquisition reporting system could be changed to present more completely the status and progress of major weapon systems and their expected operational capabilities and limitations. This review was made at the request of the Chairman, Subcommittee on Priorities and Economy in Government, Joint Economic Committee.

- GAO Report PSAD-79-46, "Need for More Accurate Weapon System Test Results to be Reported to the Congress," March 9, 1979.

  SYNOPSIS: The results of the tests and evaluations of 10 Navy and 5 Air Force weapon systems contained in the Data Sheets reviewed by GAO were found to be incomplete, misleading, and/or outdated. The three Army systems selected for review had deficiencies identified during recent operational testing; however, those test results were satisfactorily reported on the Data Sheets. The following factors have contributed to inaccurate reporting of test results to Congress: reviewers at higher levels often omit or change the Congressional Data Sheets prepared by independent test agencies; the reports are not

updated between the annual reporting periods; and the Department of Defense has not provided sufficient instructions regarding the type of information required in Congressional Data Sheets.

Recommendations include that the Secretary of Defense should improve the quality and timeliness of operational test results reported in Congressional Data Sheets through the following procedures: (1) more thorough review of Congressional Data Sheets by the Office of the Director of Test and Evaluation before submission to Congress; (2) more definitive guidelines concerning the kind of test information that is required in Congressional Data Sheets; and (3) continual evaluation and identification of new results that become available after initial Congressional Data Sheets have been submitted in order to determine if updated information should be reported to Congress.

H. MISCELLANEOUS

- GAO Report PSAD-80-61, "Implications of Highly Sophisticated Weapon Systems on Military Capabilities," June 30, 1980.

  SYNOPSIS: This report identifies the highly sophisticated nature of deployed weapons as a contributor to budgetary problems, inventory shortfalls, and a low state of readiness for certain combat categories. It points out that Defense efforts to introduce lower cost weapon systems that are more reliable, available, and maintainable have not been as successful as desired. The review, requested by the chairman of the Senate Committee on Foreign Relations, was prompted by the November 8, 1979, report on "Impediments to Reducing the Costs of Weapon Systems," PSAD-80-6.

- GAO Report PSAD-80-43, "Issues Identified in 21 Recently Published Major Weapon System Reports," June 12, 1980.

  SYNOPSIS: This report summarizes 21 major weapon system reports issued during January and February 1980. The purpose is to focus attention on the principal issues that were found to be common among several weapon programs. The report also serves as a quick reference to all major acquisition work during the past 12 months.

- GAO Report PSAD-80-6, "Impediments to Reducing the Costs of Weapon Systems," November 8, 1979.

  SNYOPSIS: GAO believes the major effects on the costs of weapons systems have resulted from: (1) attempts to deploy systems with new technology and high performance; (2) low rates of production due to budget constraints and desires to maintain active production bases as long as possible; (3) absence of price competition between contractors; (4) lack of real motivation on the part of contractors to reduce costs; (5) the impact of socioeconomic programs, government controls, and red tape; and (6) a nationwide problem of reduced research and development expenditures and lessening productivity.

  Recommendations made in this report include that the Secretary of Defense should make a comprehensive study to identify those aspects of contract administration that can be relaxed or modified in order to reduce costs and paperwork. The Secretary should also take stronger initiatives to accelerate the implementation of management policies for major weapon system acquisitions, as set forth in the Office of Management and Budget (OMB) Circular A-109. The Armed Services and Appropriations Committees should carefully examine lower cost options before approving new weapon programs. In particular, the committees should explore with senior military officials the pros and cons of larger quantities of alternative weapons versus smaller numbers of highly sophisticated and expensive systems. The Committees also should, after being satisfied that a weapon system is ready for production, consider multiyear funding in order to take advantage of more economical production practices. The Congress should take the initiative in responding to the recommendations of the Commission on Government Procurement to: (1) reexamine the full range of socioeconomic programs applied to the procurement process and the administrative practices followed in their application; and (2) raise the minimum dollar thresholds at which such programs are applied to the procurement process.

- GAO Report PSAD-79-64, "Digests of Major Weapon System Reports, Issued January and February 1979," April 25, 1979.

SYNOPSIS: This report contains unclassified digests of major weapon system reports issued during January and February 1979 and a listing of major acquisition reports issued from July 1978 through March 1979. Selected reports are included in the appropriate bibliography section.

III. WEAPON SYSTEMS REPORTS

A. AEGIS

GAO Reports

- GAO Report PSAD 80-18, "Status of the Navy's Aegis Weapon System and DDG-47 Shipbuilding Program," February 28, 1980.

- GAO Report PSAD-78-29, "Status of the Aegis Program" (SECRET), February 24, 1978.

Aviation Week and Space Technology

- "Senate Agrees to Cancel Nimitz, Equip Long Beach With Aegis," March 14, 1977, p. 22.

- Filter Center: Shipboard combat system that integrates Aegis with other weapon systems to be developed by RCA, November, 1976, p. 45.

- Industry Observer: Navy will equip surface ships with Aegis air defense system, May 3, 1976, p. 11.

Other Periodicals

- Prina, L. Edgar, "Defender of the Fleet and of the Not-So-Fleet," Sea Power, Vol. 23, April, 1980, pp. 60-62.

- ___, "Shipboard Voice Communications Being Upgraded With Integrated System," Defense Electronics, Vol. 11, July, 1979, pp. 46-48.

- Polman, Norman, "USS Norton Sound, the Newest Old Ship," US Naval Institute Proceedings, Vol. 105, April, 1979, pp. 70-83.

- Cdr. Blades, Todd, "DDG-47, Aegis on Its Way to Sea," US Naval Institute Proceedings, Vol. 105, January, 1979, pp. 101-106.

- DeVore, Charles, "Aegis: Shield of the Fleet of the 1980's," Signal, Vol. 33, October, 1978, pp. 46-47.

- Hartman, Richard, "Aegis, Navy's New AAW Weapon System," Military Electronics/ Countermeasures, Vol. 3, September, 1977, p. 14.

- Meyer, Wayne E., "Aegis, RADM," US Naval Institute Proceedings, Vol. 103, February, 1977, pp. 93-97.

- Taylor, T. W. and Wares, E. F., "A Sophisticated Test System to Support Modern Weaponry," Defense Management Journal, Vol. 13, January, 1977, pp. 16-23.

B. ALCM

GAO Reports

- GAO Report PSAD-79-19, "Air-, Sea-, and Ground-launched Cruise Missile Program Status Issues" (SECRET NOFORN), February 12, 1979.

- GAO Report PSAD-78-39, "Status of the Cruise Missile Program" (SECRET), April 28, 1978.

- GAO Report PSAD-77-36, "Status of Air and Sea Launched Cruise Missile Programs" (CONFIDENTIAL), April 26, 1977.

Aviation Week and Space Technology

- "ALCM Crash Delays Program," December 8, 1980, p. 22.

- "ALCM Test Fails," July 28, 1980, p. 17.

- "Boeing ALCM Tested," June 23, 1980, p. 18.

- "ALCM to Enter Inventory," June 16, 1980, p. 76.

- "Boeing Gears for Production," May 12, 1980, p. 63.

- "Software Key to ALCM Choice," March 31, 1980, p. 18.

- "ALCM Contractor to Be Selected," February 18, 1980, p. 22.

- "Weather Could Prevent Flyoff," December 3, 1979, p. 22.

- "Cruise Missiles in Flight Tests," October 15, 1979, p. 74.

- "ALCM Test Terminated," October 15, 1979, p. 22.
- "Boeing Probes Crash," August 13, 1979, p. 18.
- "Cruise Missile Flyoff Begins," July 30, 1979, p. 14.
- "Disperser Modified for ALCM," April 2, 1979, p. 66.
- "First ALCM-B Test Missile in Fabrication," July 24, 1978, p. 61.
- "Cruise Missile Launch Concepts Detailed," June 12, 1978, p. 56.
- "USAF Plans Radar Guidance," May 15, 1978, p. 17.
- "ALCM Undergoing Tests," February 20, 1978, p. 38.
- "Tomahawk, ALCM Survivability Tests," December 12, 1977, p. 17.
- "New Air-Launched Cruise Missile Mode Shown," December 5, 1977, p. 17.
- "Cruise Missile Flyoff Planned," July 18, 1977, p. 12.
- "Competitive Cruise Missile Work Urged," July 11, 1977, p. 22.
- "Boeing Set for Cruise Missile Speedup," July 11, 1977, p. 19.
- "USAF Pushes Long-Range Cruise Missile Version," April 18, 1977, p. 19.
- "Boeing USAF Missile Effort Set," March 14, 1977, p. 23.
- "Cruise Missile Order," January 24, 1977, p. 17.
- "USAF May Accelerate Missile Programs," October 11, 1976, p. 14.
- "ALCM in Flight Tests," July 12, 1976, p. 19.
- "WRU: DOD to Study 2 Strategic Weapons Plans," May 31, 1976, p. 11.
- "ALCM Maneuverability Shows," May 24, 1976, p. 25.

Other Periodicals

- Meyer, Deborah G., "Boeing Lands ALCM Contract," Armed Forces Journal International, Vol. 117, May 1980, p. 17.
- Maj. Robinson, Gerald L., "ALCM Construction Program," Air Force Engineering and Services Quarterly, Vol. 21, May 1980, pp. 8-30.
- ___, "New Missile (ALCM) Picked," Air Force Times, Vol. 40, April 7, 1980, p. 7.
- Meyer, Deborah G., "ALCM Competitors Complete Flight Tests," Armed Forces Journal International, Vol. 117, March 1980, p. 31.
- Lopez, Roman, "ALCM--Who Will Build the ALCM," Interavia, Vol. 35, March 1980, pp. 206-208.
- ___, "Cruise Missile Development: The Competition Is Fierce Everywhere," Government Executives, Vol. 12, January 1980, pp. 22-24.
- Maj. Tucker, Charles G., "The ALCM Decathlon," Air Force Magazine, Vol. 62, June 1979, pp. 67-71.
- Berry, Jr., F. Clifton, "Cruise Missile Developments in the Aftermath of SALT II," Interavia, Vol. 34, August 1979, pp. 760-764.
- Lamb, J.E., "Cruise Missile - Superweapon," National Defense, Vol. 62, Sept-Oct 1977, pp. 120-122.
- Brown, Harold, "The B-1 Cruise Missile Decision," Commanders Digest, Vol. 20, September 15, 1977, Entire Issue.
- Clements, Williams P., "View From the Top (Deputy Secretary of Defense)," Countermeasures, Vol. 3, Jan 1977, p. 6.
- Berry, F. Clifton, "ALCM Flying High--Is It Doomed?" Armed Forces Journal International, Vol. 114, Oct 1976, p. 22

C. EF-111A

GAO Reports

- GAO Report PSAD-80-16, "Lack of Full Disclosure of EF-111A Program Cost," December 4, 1979.

- GAO Report PSAD-80-71, "Evaluation of EF-111A Extended Development and Full-Scale Production Decision," August 27, 1980.
- GAO Report PSAD-79-5, "Assessment of Joint DT&E/IOT&E Results of EF-111A" (SECRET), November 3, 1978.
- GAO Report PSAD-79-74, "Recommendations for Additional Testing Prior to Full-Scale Production of EF-111A," April 25, 1979.
- GAO Report PSAD-78-130, "Recommendations for Improving Testing of EF-111A Tactical Jamming System," June 30, 1978.
- GAO Report PSAD 78-32, "Status of the Air Force's EF-111A Tactical Jamming System" (SECRET), March 23, 1978.
- GAO Report PSAD-77-25, "Status of the EF-111A Electronic Warfare Aircraft Program" (CONFIDENTIAL), February 3, 1977.

Aviation Week and Space Technology

- "EF-111A Conversion to Start," August 11, 1980, p. 25.
- "EF-111A Exceeds Goals," October 29, 1979, p. 68.
- "EF-111A Costs Cause Concerns," July 17, 1978, p. 15.
- "USAF Improves ALQ-99," February 6, 1978, p. 97.
- "EF-111A Geared for Electronic Thicket," February 6, 1978, p. 93.
- "EF-111 Electronic Warfare Aircraft," February 6, 1978, p. 56.
- "ECM-11 in Tests," March 28, 1977, p. 58.

Other Periodicals

- Capt. Stoehrmann, Kenneth C., "New Roles for TAC's F-111," Air Force Magazine, Vol. 63, February 1980, pp. 46-51.
- Ulsamer, Edgar, "EF-111A (Tactical Jamming System) USAF's Supersonic Electronic Shield," Air Force Magazine, Vol. 60, August 1977, pp. 54-59.

. Sundaran, G.S., "E-W Dedicated Aircraft : the EA-6B and EF-111A," International Defense Review 10, February 1977, pp. 54-58.

D. MCLWG

GAO Reports

. GAO Report PSAD-78-122, "An Assessment of the Major Caliber Lightweight Gun System" (SECRET), August 4, 1978.

. GAO Report PSAD-77-4, "Operational Testing of the Major Caliber Lightweight Gun," November 5, 1976.

E. PATRIOT

GAO Reports

. GAO Report PSAD-78-15, "Satus of the Patriot Program" (SECRET).

. GAO Report PSAD-76-89, "SAM-D Missile," February 17, 1976.

Aviation Week and Space Technology

. "Brown Approves Limited Patriot Production," September 22, 1980, p. 22.

. "Germans Weighing Alternative Patriot Acquisition Technique," August 29, 1977, p. 47.

. Filter Center: System capable of flying target drones in formation for SAM-D tests to be delivered, May 31, 1976, p. 51.

. Industry Observer: Marines decide SAM-D is too large for tactical assault mission, May 3, 1976, p. 11.

. "Army Presses New Missile," March 15, 1976, p. 23.

Other Periodicals

. Meller, Rudi, "The U. S. Army's Mobile Air-Defense Missile System Nears Production," International Defense Review 13, Vol. 4, December 1980, pp. 495-499.

. Power, Tom, "Drones Take on White Sands Missiles," Air Force Times, Vol. 40, May 19, 1980, p. 31.

- Lt. Col. Naegele, Eugene L.," Automated Operation With Human Control, A Key Feature of Patriot," Air Defense Magazine, July-September 1978, pp. 17-23.

- ___, "DOD Favors Surface-to-Surface Version of Patriot for Assault Breaker Role" Armed Forces Journal International Vol. 115, August 1978, p. 13.

- ___, "Patriot," Signal, Vol. 32, April 1978, p. 71.

- Lt. Col. Neagele, Eugene L., "Track via Missile Guidance," Air Defense Magazine, January-March 1978, pp. 12-16.

- Lt. Col. Neagele, Eugene L. "Multifunction Phased Array Radar, a Key Feature of Patriot," Air Defense Magazine, October-December 1977, pp. 37-40.

- Capt. Anderson, R.E., "Patriot - Continuation of a Proud Heritage, Air Defense Magazine, April-June 1977, pp. 32-36.

F. ROLAND

GAO Reports

- GAO Report PSAD-79-91, "Evaluation of the Decision to Begin Production of the Roland Missile System" (SECRET), June 29, 1979. Unclassified version PSAD-79-100, August 17, 1979.

- GAO Report PSAD-79-28, "How Ready Is the Army's Roland Missile for Production?" February 14, 1979.

- GAO Report PSAD-78-128, "Status Report on the Roland Missile Program," July 7, 1978.

- GAO Report PSAD-78-16, "Issues To Be Resolved Before Committing the Army's Roland Missile System to Production," March 10, 1978.

- GAO Report PSAD-77-38, "Status of the Roland and Chaparral Missile Programs" (CONFIDENTIAL), February 28, 1977.

- GAO Report PSAD-76-116, "Roland II/Chaparral/AAA/ LOFADS Systems," March 17, 1976.

Aviation Week and Space Technology

. "U.S. Falters on Commitments to NATO," November 10, 1980, p. 35.

. "Tornado Costs Delay Roland 2 Buy," October 20, 1980, p. 113.

. "Roland Demonstration Tests Scheduled," August 18, 1980, p. 81.

. "Army Limits Roland Production," January 14, 1980, p. 66.

. "Roland Leaves Launch Tube to Intercept Aerial Target," June 11, 1979, p. 86.

. "Funding Approval for Roland May Spur Senate, House Clash," June 4, 1979, p. 21.

. "Roland Proves Feasibility of Technology Transfer," September 11, 1978, p. 114.

. "Boeing Tests Roland," May 22, 1978, p. 25.

. "Roland Tests Start," February 20, 1978, p. 18.

. "Roland Effort Tests Technology Transfer," November 14, 1977, p. 18.

. "Roland 2 Nearing U.S. Production," January 17, 1977, p. 57.

. "Roland Service Introduction Nearly," May 3, 1976, p. 55.

. "Issue of Roland Sale by U.S. Unresolved," January 12, 1976, p. 9.

Other Periodicals

. Robins, Yves, "Roland: Answer to the Low-Level Air Threat," NATO's Fifteen Nations, Vol. 25, February-March 1980, pp. 60-66.

. ___, "Rapier Buy, Roland Out?" Armed Forces Journal International, Vol. 117, September 1979, p. 10.

. ___, "Roland Launched," Military Review, Vol. 59, September 1979, p. 76.

- Geddes, J. Philip, "Technology Transfer and the U.S. Roland Missile," Interavia, Vol. 33, November 1978, pp. 1095-1099.

- Shohat, Murry, "Production Greased for Roland," Military Electronics/Countermeasures, Vol. 4, September 1978, p. 28.

- ___, "Roland: Exchange of Expertise," National Defense, Vol. 62, May-June 1978, pp. 549-550.

- Brig. Gen. Ragano, Frank P., "U.S. Roland: A Milestone in Technology Transfer," Defense Management Journal, Vol. 14, May 1978, pp. 3-10.

- ___, "Roland: Army's Leading Case-In-Point," Government Executive, Vol. 10, March 1978, p. 23.

- ___, "U.S. Roland: A Comprehensive Update," Air Defense Magazine, January-March 1978, pp. 37-39.

- Richardson, John H., "Roland, A Technology Transfer Program," Defense Systems Management Review, Vol. 1, Summer 1977, pp. 13-17.

- Brig. Gen Ragano, Frank P., "U.S. Roland--A Giant Step Toward Weapon Commonality," Defense Systems Management Review, Vol. 1, Summer 1977, pp. 9-12.

- Hewish, Mark, "Weapon Systems, Roland," U.S. Naval Institute Proceedings, Vol. 103, June 1977, p. 122.

- ___, "U.S. Roland Progress, More Than a Missile," Air Defense Magazine, April-June 1977, pp. 46-47.

- Walsh, Bill, "Roland, the U. S. Army's New International Missile System" Countermeasures, Vol. 3, March 1977, p. 22.

- Brig. Gen. Ragano, Frank P., "View From the Top," Countermeasures, Vol. 3, March 1977, p. 6.

APPENDIX B

INTERVIEW GUIDE

ADEQUACY OF T&E DATA

- Interview Guide -

INTRODUCTION

Booz, Allen is conducting a research study for the Defense System Management School regarding the adequacy of test information provided at the time of program milestone decisions. It has been observed that in numerous cases the DSARC has rejected the services' proposal for moving into the next phase or has attached qualifications involving additional T&E because the data presented to them is "inadequate." To date no one has been able to precisely define what constitutes adequacy but we are trying to characterize the scenarios in which this phenomena occurs and to understand the reasons for it. We hope to be able to make recommendations—perhaps involving fundamental changes to acqusition policy—which will alleviate it in the future. Your program has been selected for detailed review and specifically we are looking at the DSARC that was held on _____.

1. As background, we are trying to establish a decision profile for this system. Can you describe the program plan as of the previous DSARC? The position of the Service going into this DSARC? The decision resulting from this DSARC?

2. Was there an approved TEMP at the time of the previous milestone decision? or shortly thereafter? (reference) If there wasn't, why not?

3. Could you comment on your view of the utility of the TEMP. (Probe for sense of contract with OSD vs useless appendage.)

4. If there was a TEMP, in your judgement was it well conceived? For example, did it:

- Address operational suitability as well as technical performance
- Propose evaluation in terms of actual threat in a realistic environment
- Require typical military personnel for operation and maintenance
- Provide for adequate test resources.

5. Was the test program described in the TEMP (or otherwise originally planned) carried out? If not, why not? For example:

- Lack of money
- Slips due to technical problems
- Lack of proper test facility/instrumentation
- Other reasons.

6. Could you describe the test program itself, including:

   - Were requirements well-defined before testing started?
   - Did requirements change after testing began?
   - Were adequate resources available? (test items, facilities, etc.)
   - Was test cost a factor? or did testing get squeezed by other overruns in the program?
   - Other comments regarding the conduct of the DT/OT phase.

7. What did the test results show? What was the position of the independent evaluator?

- Issues relate primarily to DT or OT
- Failure to meet thresholds
- Was software an issue? How?

8. If there was a case of "inadequate data," what is your view of why it occurred?

9. One frequently heard complaint is that the items being tested in IOT&E are not sufficiently mature or representative of the production item to yeild valid data DOD policy permits limited production of items for operational testing prior to Milestone III. Why wasn't this done?

10. Why did the Service go to the DSARC anyway? (With inadequate data.)

11. What could have been done differently, how could the situation have been improved?

12. In your opinion, did other external factors influence the decision of the DSARC? For example,

- Affordability (either in near or far term)
- Other program failures/controversies
- Recent studies (e.g. GAO)
- Congressional actions
- Dominant personalities within OSD (with pet theories)
- National situation.

13. Do you have any comments on reasonable expectation of DSARC principles?

- How much is enough? (Confidence, sample size, etc.) Can you ever have too much?
- How should OSD integrate the total body of data and weigh with other aspects (e.g. urgency)?
- What should the role be of simulations and/or analysis as a supplement to empirical tests?
- How should one treat R&M growth?
- What was the impact of the DDT&E assessment?
- Was there an awareness of the cost (resource implications) of decisions to delay and the impact on contracts the POM process, etc.
- Did OSD hold up there side of the "contract" regarding goals and thresholds, or was it a new ball game?

APPENDIX C

OVERVIEW OF CHANGES IN ACQUISITION POLICY

OVERVIEW OF CHANGES IN ACQUISITION POLICY

This Appendix contains a brief overview of some of the key studies and other events since 1970 that influenced acquisition policy.* Specifically, these include:

- The memoranda to the Secretaries of the military departments by Deputy Secretary of Defense David Packard in 1969 and 1970
- The final report in 1972 of the Commission on Government Procurement (created by Congress in 1969) which eventually led to the creation of the Office of Federal Procurement Policy and the publication of the Office of Management and Budget (OMB) Circular A-109
- The formation of the Acquisition Advisory Group (AAG) by Deputy Secretary of Defense William P. Clements in 1975
- The designation of the first Defense Acquisition Executive in 1976
- The Defense Science Board Summer Study, Report of the Acquisition Cycle Task Force, published in March 1978
- The Defense Resource Management Study, known informally as the Rice Report, published in February 1979 which led to the creation of the Defense Resources Board.

(1) Packard Initiatives

The Packard memos of 1969 and 1970 began the modern era of acquisition reform and established much of the framework for the acquisition process as it exists today. Packard saw that the decentralization of operations, which had been a goal of the previous administration (along with centralized decision authority and planning), had not occurred. He believed that OSD should be removed from the detailed

---

* A comprehensive overview of the development of the defense system acquisition process is given in "The Maturing of the DOD Acquisition Process," by David D. Acker in the Defense Systems Management Review, Volume 3, Number 3, Summer 1980.

direction of individual defense programs and, instead, that OSD should assume a role of monitor and decision-maker at milestones associated with major system. Packard offered this change as the best hope of improving the management of the defense system acquisition process and of gaining control of system acquisition costs. Among Packard's initiatives were the following:

- Use of demonstrated performance as the primary measure of a system's readiness to proceed from one phase of the acquisition cycle to the next. This essentially terminated the Total Package Procurement concept which had been used before.

- Institution of the Development Concept Paper (DCP, later the Decision Coordinating Paper) as a "contract" between OSD and the Services which established the program objectives, including technical requirements, schedule and cost, and the thresholds which, if breached, would trigger a review of the entire program.

- Establishment of the DSARC as an advisory body to the Secretary of Defense on the status and readiness of major system to progress to the next acquisition phase.

The DSARC as originally conceived was intended to complement the DCP approval process by providing a means for coordinated review and evaluation by the senior OSD staff. Thus, the original purpose of the DSARC was as much to inform the staff of the status of designated programs as it was to have them sit in judgment of it. This was consistent with Packard's desire to disengage OSD from the day-to-day operation of individual programs.

Other changes by Packard included strict avoidance of unnecessary concurrency between development and production; emphasis on hardware competition and prototypes, rather than analysis ("paper studies"), as a means of judging the value of, and selecting between, alternative weapon systems; and clarification of responsibilities of OSD, the military departments, and the program manager. Also, Packard addressed the problem of cost growth in defense systems. One of his recommendations eventually led to the formation in 1972 of the Cost Analysis Improvement Group (CAIG) as another input to the DSARC milestone review process.

In the 1969 time frame, the President and the Secretary of Defense chartered a Blue Ribbon Defense Panel to study the entire organization, structure, and operation of DOD. Many of the resulting findings were directly related to defense system acquisition. Secretary Packard used the Blue Ribbon Defense Panel report and his own experience to formulate a cohesive statement of DOD acquisition policy in the form of the first DOD Directive 5000.1, Acquisition of Major Systems, in July 1971.

(2) Commission on Government Procurement

During this same period, Congress was exhibiting increased dissatisfaction with the entire Federal procurement process. In November 1969, Public Law 91-729 was enacted which established the Commission on Government Procurement (COGP) to make recommendations on improving the effectiveness and efficiency of Federal procurement practices. After nearly 3 years and an exhaustive series of hearings, the Commission published its report in December 1972. It made 149 sweeping recommendations, 12 of which concerned the system acquisition process. These recommendations were focused primarily on the "front end" of the process where the Commission found that most problems began. The recommendations were grouped in five major categories:

- Establishing needs and goals
- Exploring alternative systems
- Choosing a preferred system
- System implementation
- Organization, management, and personnel.

The recommendations were generally well received and independently supported by the DOD Acquisition Advisory Group (described below). In April 1976, the Office of Federal Procurement Policy (which was itself an outgrowth of the COGP, having been established by Congress in late 1974 pursuant to P.L. 93-400) issued OMB Circular A-109 which embodied the acquisition philosophy that had emerged. A-109 articulated a system approach to acquisition which combined problem solving logic, increased flexibility, and several fundamental management and business principles. The management principles involved:

- Establishing clear responsibilities and authorities for the program manager

- Generating a tailored acquisition strategy early in the program to guide the entire effort

- Requiring approval by the agency head for continuing the program at key decision points

- Encouraging creative new ideas by avoiding government specified solutions to defined problems.

The business principles included:

- Creating and maintaining competition as far along the process as is economically beneficial

- Continuing contractor work without interruption as the Government goes about its decisionmaking

- Recognizing that the Government must pay for the benefits that result from increased competition.

The DOD has gradually adopted the precepts of A-109. Revisions to DOD Directive 5000.1 and DOD Instruction 5000.2 have incorporated a new milestone 0 and emphasized the front-end mission analysis by requiring a mission element need statement (MENS). While there have been critics who attribute increased time and cost to A-109 policies, the present philosophy appears to be that the policies themselves are sound, and that improvements are to be made by better implementation of the fundamental principles.

(3) Acquisition Advisory Group

In 1975, Secretary Clements created the AAG to review the separate recommendations of the three Services on improving the acquisition process. The AAG found that the existing policies as stated in DOD Directive 5000.1 were sound but that improvements were needed in the way the policies were implemented by both OSD and the Services. In particular, the AAG stressed decentralization of the management of individual programs and a stricter interpretation of the DSARC role in providing advice to OSD on designated major systems only at formal milestone points. By and large, the recommendations of the AAG were consistent with those of the COGP and supported the management thinking that led to A-109.

(4) Defense Acquisition Executive

Responding to A-109, the DoD appointed its first Defense Acquisition Executive (DAE) in 1976. The DAE was to be the primary advisor to the Secretary of Defense on matters relating to the acquisition of defense systems. The significance of the appointment lay in the fact that previously there was a strong distinction between the R&D community, which was the responsibility of the DDR&E, and the production community, primarily the purview of the Assistant Secretary of Defense (Installations and Logistics) (ASD(I&L)). Because the two communities viewed the process differently, participated to different degrees at different times, and generally did business in different ways, there was a cultural gap between them which compounded the problems of transitioning between R&D and production. The DAE's responsibility was to integrate and unify the management process; however, there are some indications that vestiges of divided cognizance still exist.

(5) Acquisition Task Force

A somewhat contrary note was sounded in March 1978 with the publication of the report of the Acquisition Cycle Task Force of the Defense Science Board. The Task Force examined a number of major systems acquisition case histories, changes in acquisition policy over the past two decades, and the actual workings of the program advocacy and budgetary processes. They concluded that, over the past 15 to 20 years, the acquisition process--from initial program concept to initial operational capability--had increased in length by approximately 5 years on the average. Another conclusion was that most of this increase had been in the front end of the process, i.e., from the time of initial concept studies to the start of Full-Scale Development (FSD) and at the back end from the completion of FSD to the achievement of a full operational capability.

The Task Force also concluded that full-scale prototyping often introduced costly and unnecessary program delays, that the policy of "no concurrency" was being applied too rigidly and was inefficient and costly in many cases, and that sequential and separate testing of systems by developers and users (instead of the independent evaluation of jointly obtained test data) often added unnecessary expense and time. It was also concluded that the practice of incremental

decisionmaking--by requiring the achievement of a consensus between the advocates and the antagonists at each milestone--was a cumbersome process that tended to prolong programs so much that the threat, the players, and even the concepts of warfare often changed during the acquisition cycle. This resulted in canceling programs and starting new ones to meet the threat with the concomitant waste of limited national resources.

Finally, the Task Force expressed concern that without very careful attention to the implementation of OMB Circular A-109, procedures could be institutionalized that would significantly increase the total length of the acquisition process without achieving the intended degree of program flexibility and stability in acquisition policy.

The DSB made recommendations regarding affordability of new weapon systems, flexibility of the acquisition process, tailoring of phases and milestones, limiting phototyping, increasing concurrency when warranted, combining development and operational testing, and coupling the DSARC and PPBS. In addition, it suggested that product improvement of existing system be considered as an alternative to new system development whenever possible.

Some acquisition reformists felt that the DSB report overly favored expediency as an operating mode. Nevertheless, most of their recommendations were reflected in the March 1980 revisions to 5000.1 and 5000.2. As a result, there was a shift in the swing of the pendulum away from excess caution toward greater risk-taking.

One of the more important changes of these latest revisions to 5000.1 and 5000.2 in terms of understanding the decision process is that milestones 0 and II were given important new meanings. Milestone 0 now reflects a commitment on the part of the Secretary of Defense to "satisfy the need" and milestone II his intention to "deploy the system." The significance of the latter is that it is now officially acknowledged that the major commitment to the program is really to be made at milestone II and not at milestone III as

---

* Exhibit III-1, taken from the Booz, Allen wall chart, "Acquisition Process for Major Defense Systems," February 1981, discusses the full range of policy changes that were thus introduced.

EXHIBIT C-1
March 1980 Changes in Policy

MARCH 1980 CHANGES IN POLICY

On March 19, 1980, the Department of Defense extensively revised the articulation of its top-level policy and procedural framework for the acquisition of major systems. DOD Directives 5000.1 and 5000.2 – last revised and issued on January 18, 1977 for the purpose of implementing OMB Circular A-109 – were cancelled and replaced by DOD Directive 5000.1 "Major System Acquisitions" and DOD Instruction 5000.2, "Major System Acquisition Procedures." The following is a listing of some of the more important changes in policy (including emphasis) and in procedures as reflected in the current revisions:

- A-109. There is a far stronger commitment to embracing the fundamental precepts of OMB Circular A-109. The circular is now enclosed with DOD Directive 5000.1 (rather than referenced) and a number of its key objectives are emphasized throughout both 5000.1 and 5000.2 which are established as having the highest precedence over all other DOD acquisition-related issuances.

- AFFORDABILITY. The concept of "affordability" – the overall reconciliation of resources and needs – has been emphasized and made a major determinant in the decision-making process. At each of the decision milestones the DOD Component must provide assurance that resources are available or can be programmed to execute the program in an efficient and effective manner.

- PPBS INTERFACE. The persistent problem of better interface between acquisition milestone decisions and annual PPBS overall resource allocation decisions has been given increased emphasis. In particular, the USDRE, ASD(C) and the ASD (PA&E) are given bridging responsibility for reviewing affordability issues (e.g., cost, priority, and availability of fiscal and manpower resources) in the context of the PPBS process.

- MILESTONE DECISION COMMITMENT. Decisions by the SECDEF at Milestones 0 and II have been given important new meanings. Previously, the Milestone 0 decision represented a commitment only to identify and explore alternative solutions and the Milestone II decision represented a commitment only to continue the program through the engineering development phase. The decisions now reflect a commitment on the part of the SECDEF to "satisfy the need" (Milestone 0) and to "deploy the system" (Milestone II).

- MAJOR SYSTEMS DESIGNATION. Rules for designating system as "major" have been relaxed in favor of more pragmatic guidelines. Rules remain for determining when a MENS is required ($100M RDT&E or $500M procurement) but the resulting system may or may not be designated as major. This is consistent with the often stated objective of reducing the number of systems going through the formal DSARC process.

- TIMELINESS, TAILORING AND CONCURRENCY. The concept of "tailoring" the acquisition process, as described in OMB Circular A-109, is not new; however, the concept of planned "concurrency," expressly proscribed during the Packard era, is cautiously reintroduced as a means of minimizing acquisition time when technical, cost and supportability risks are low or when urgency to counter a threat is paramount. Examples cited include: increasing funding, overlapping, combining, or omitting acquisition phases or overlapping or combining development T&E with operational T&E. Also, the concept of accepting system performance growth after deployment is included as a means of achieving improved responsiveness to the threat.

MARCH 1980 CHANGES IN POLICY (con't)

- **DEFENSE SYSTEMS ACQUISITION REVIEW COUNCIL.** DSARC membership has been reconstituted: the Under Secretary of Defense for Policy and the Chairman, Joint Chiefs of Staff have been added as permanent members; the ASD (C^3I) has been made a principal advisor, (reflecting an organizational change and replacing the ASD(I) and Director, Telecommunications who were previously members); six OSD officials have been added as principal advisors; and the Service Secretaries whose programs are under review have been eliminated as designated participants. Further, milestone decision making and review procedures have been refined, OSD responsibilities clarified and new information and documentation requirements (such as the IPS, MRF and SDDM) have been imposed. In order to discipline the process, a rigid 6-month timetable for staffing has been added. See DSARC Milestone Planning Schedule at bottom.

- **INTERACTIVE THREAT ANALYSIS.** The key interaction between intelligence collection and weapon system performance and design has for the first time formally been made a part of the acquisition process. A continuing "interactive analysis" – a study of the system-threat interaction – is now required with the first submission to be completed before Milestone I. The result of this process is a set of Critical Intelligence Parameters (CIPs) which are controlled and updated throughout the life of the program. DOD Components are required to identify CIPs in order to guide collection activities, and DIA is required to validate threat data, review CIPs output and report findings and conclusions to the DAE prior to DSARC meetings.

- **ACQUISITION STRATEGY FOR C^3 SYSTEMS.** Certain command and control systems have been formally recognized as differing from other weapon systems and as having major characteristics which require special management procedures. For such systems, an acquisition strategy is suggested which allows early implementation and field evaluation of a prototype using existing commercial or military hardware or software. This permits design and testing in an evolutionary manner. In such a strategy the Concept Exploration and Demonstration and Validation phases normally are combined and, in some cases, procurement for initial fielding may proceed without a Full-Scale Development phase. Since tailoring of the acquisition process is an inherent part of developing an acquisition strategy, it is not clear why C^3 systems have been separately highlighted. A more significant distinction is that in C^3 systems the designated User is now tasked to perform operational testing (rather than an independent operational tester) and has been assigned the major responsibility for the Demonstration and Validation phase (rather than the program manager).

- **NATO RSI.** Consideration of NATO doctrine, threat assessment and mission needs and the standardization or interoperability of U.S. equipment with that of other members of NATO (as required by the "Culver-Nunn Amendments") remains a tenet of the directives. However, the revisions to DODI 5000.2 significantly increase opportunities for NATO country participation in the U.S. acquisition process by requiring that consideration be given to all existing and developmental NATO systems as an alternative to U.S. development, and that, under some conditions, NATO member contractors be solicited for bids and proposals on U.S. systems and components. In addition, international codevelopment, coproduction, and cooperative testing is emphasized.

- **MANPOWER, TRAINING AND LOGISTICS SUPPORTABILITY.** The current Directive and Instruction further emphasize that logistics supportability is a design requirement as important as cost, schedule and performance. They direct that manpower and personnel factors, such as numbers, occupations, and skill levels required, be included as considerations and constraints in system design and that early in the acquisition process manpower requirements be subjected to trade-offs with system characteristics and support concepts in order to achieve the most cost-effective balance between acquisition and ownership costs and system effectiveness.

- **ROLE OF GOVERNMENT LABORATORIES.** The formal role of the laboratories in the acquisition process has been diminished. Formerly, the program manager was required to engage a DOD laboratory for the purpose of evaluating risks and preparing a Technology Assessment Annex to the DCP. While formal risk assessment remains an inherent part of the acquisition process, the former requirement has been eliminated and the role of the laboratories is left to the services.

generally believed. The DSB called for establishment of milestone III as the decision for rate production, reflecting their concern that programs were being unnecessarily stretched-out by limited production phases. Actually, the fact that milestone II includes the authority for limited production of hardware for operational test and evaluation is stated in A-109 and has been DOD policy since 1977, but it is not commonly understood or followed.

(6) Defense Resource Management Study

When President Carter and Secretary of Defense Brown took office in 1977, they commissioned the Defense Resource Management Study (DRMS). This study was to address the resource allocation process and four of its specific functions including weapon systems acquisition. (Other functions were logistic support, enlisted personnel management, and military health care.) The DRMS proposed significant changes in the PPBS including the establishment of a Defense Resources Board (DRB) to manage a combined program/budget review during each POM cycle. Secretary Brown established the DRB in March of 1979, and subsequently changed the membership of the DSARC to coincide with that of the DRB. Thus, he attempted to achieve greater coupling between the DSARC and PPBS processes.

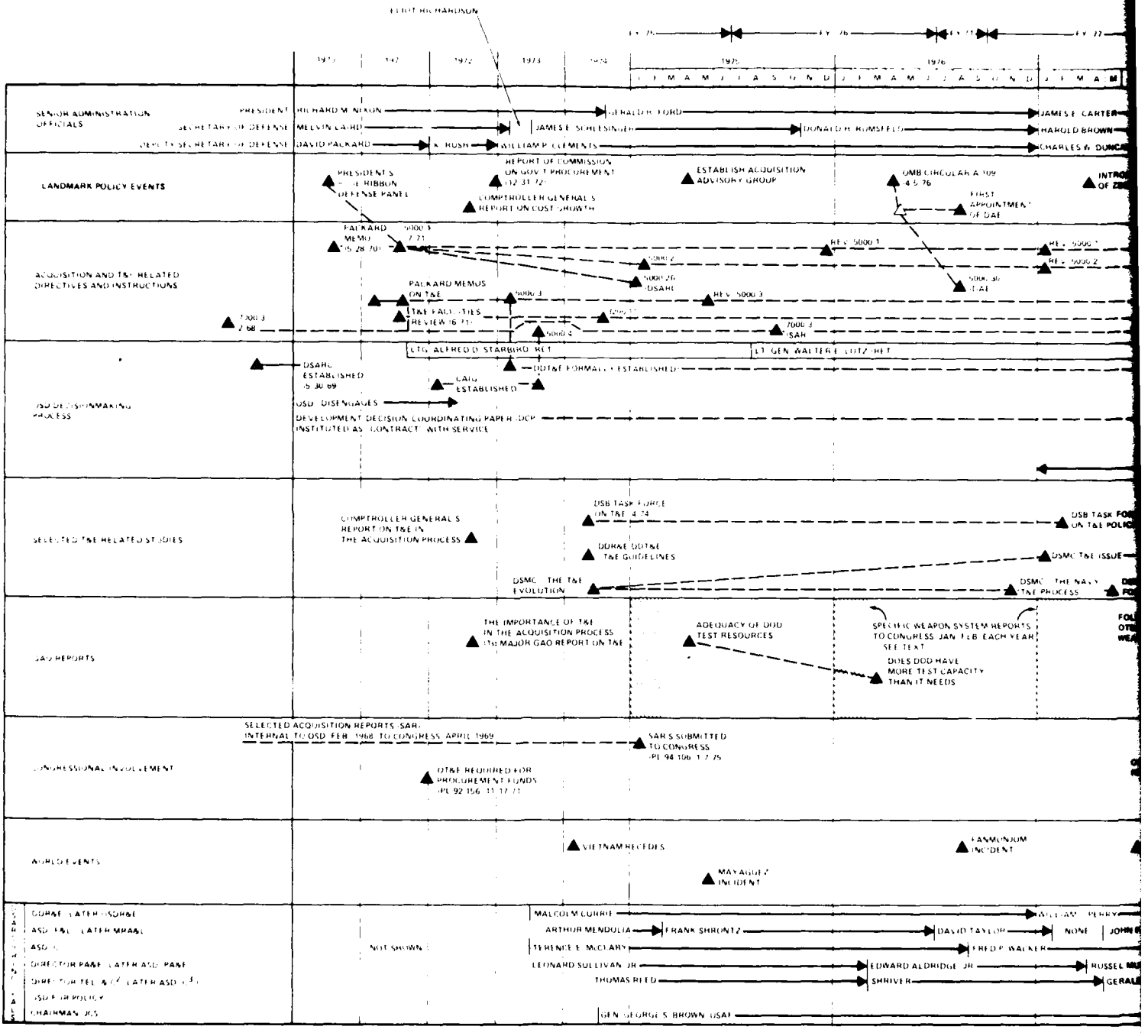
With respect to acquisition policy, the DRMS offered suggestions on the implementation of A-109, stressed competition in the acquisition process, and observed that "Troublesome and costly problems often arise from the premature commitment of systems to high-rate production." The DRMS recommended the delay of high-rate production until both technical adequacy and operational suitability (including reliability, supportability, and readiness characteristics) were demonstrated.

This DRMS recommendation is highlighted because it reflected a growing concern during the later 1970s regarding deficiencies in the operational suitability of U.S. military equipment. Championed by new personnel within OSD, operational suitability increasingly became a major topic in DSARC reviews. In January 1980 the Office of the Under Secretary of Defense, Research and Engineering (OUSDR&E)/DDT&E and Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics) (OASD(MRA&L)) jointly published the Operational Suitability Verification Study which attempted to improve operational suitability evaluation, particularly as it applied to the test community and users of test data.

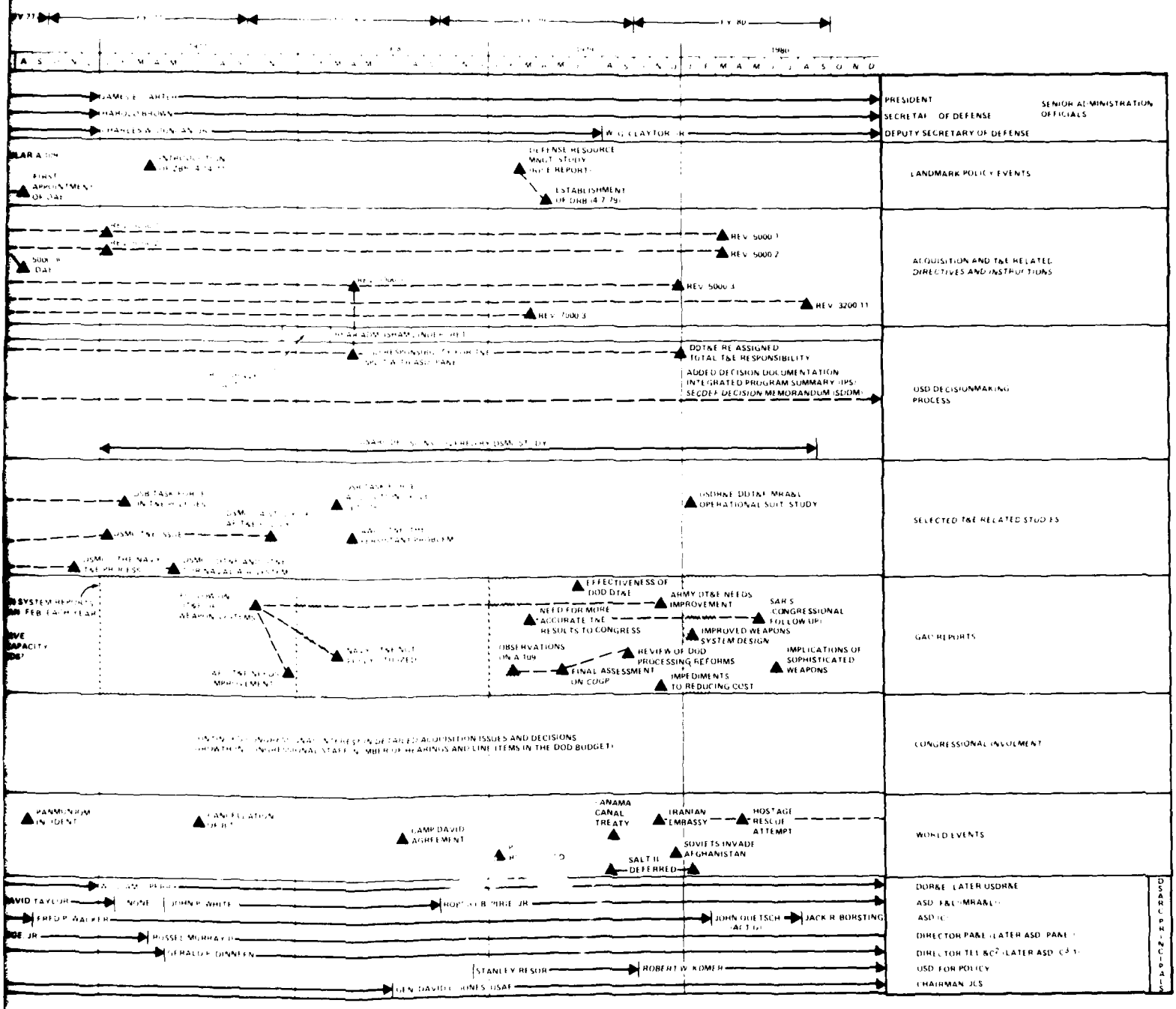
APPENDIX D

DICHRONIC CHART

HISTORICAL PERSPEC



PERSPECTIVE



APPENDIX E

CHI SQUARE TEST OF INDEPENDENCE

CHI SQUARE TEST OF INDEPENDENCE

One way to test the statistical relationship between the type of DSARC review and the likelihood of unconditional approval is by application of the chi square test of independence. Shown below is a 2 x 2 matrix, commonly referred to as a contingency table, which describes the data previously discussed.

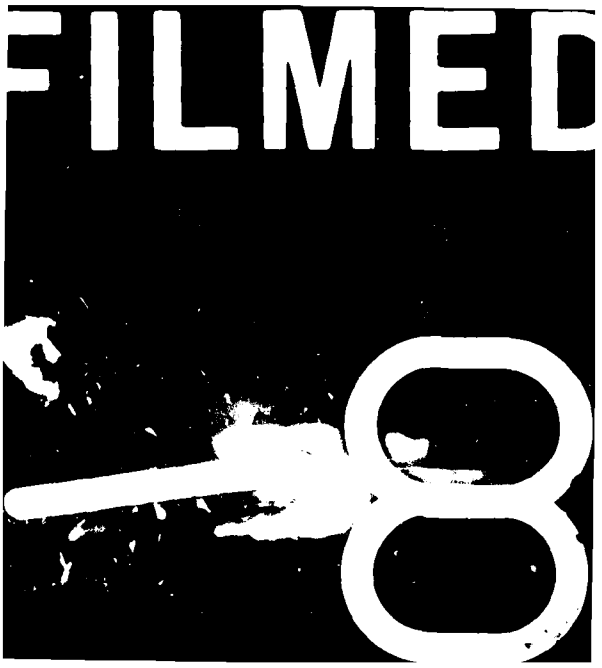
	UNCONDITIONAL APPROVAL	ALL OTHER DECISIONS	TOTAL
DSARC II	6	10	16
DSARC III	4	18	22
TOTAL	10	28	38

The chi square procedure permits a judgement as to whether there is an interaction between the type of review and the type of decision. In this case the null hypothesis, H_0, is that there is no relationship between the two; i.e., programs are just as likely to receive unconditional approval at Milestone II as they are at Milestone III. This hypothesis is rejected only if: $\chi^2_0 > \chi^2_{\alpha, \text{d.f.}}$ where:

$$\chi^2_0 = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

and O_{ij} are the observed frequency for the ith row and jth column, and E_{ij} the expected frequency (based on pure chance) for the same row and column. Alpha (α) is the confidence limit of the test, and d.f. is the number of degrees of freedom, which in this case is one.

The test statistic χ^2 is computed to be 1.799, and since $\chi^2_{.05, 1} = 3.841$ we cannot reject the hypothesis of independence.


FILMED